CPT, Modified Gravity and the Baryon Asymmetry of the Universe

Maria Pestana da Luz Pereira Ramos

Mestrado em Física (Especialização Teórica)

Departamento de Física e Astronomia

2017

Orientador

Jorge Tiago Almeida Páramos, Professor Auxiliar Convidado,
Faculdade de Ciências da Universidade do Porto

Todas as correções determinadas pelo júri, e só essas, foram efetuadas.

O Presidente do Júri,

Porto, ______/______/_________

"Tyger Tyger, burning bright,
In the forests of the night;
What immortal hand or eye,
Could frame thy fearful symmetry?"

– William Blake, *The Tyger*

Acknowledgements

First of all, I want to thank my supervisor, Jorge Páramos, for allowing me to study a subject I am so interested in. Thank you for always being available to discuss my doubts while teaching me that every professor should offer his students (a lot of) caffeine, if they choose to follow physics.

My gratitude to professor Orfeu Bertolami who has made a lot of research concerning the topics of this thesis and who was readily available to discuss them with me. I would like also to thank the other professors who were able to arouse my curiosity and clarify my ideas.

In these five years, I had the fortune to be in a room with people that I admire, who not only shared their complete and deep vision of physics with me, but who also are kind human beings. For that, I am truly grateful to João Pedro Pires, João Guerra, António Antunes, Simão João and Bruno Madureira.

Last but not least, I thank my family for the endless patience, love and for giving me the opportunity to do whatever I wanted, and my best friend, Beatriz Álvares Ribeiro, for the daily encouragement and optimism throughout this period.

UNIVERSIDADE DO PORTO

Abstract

Departamento de Física e Astronomia
Faculdade de Ciências da Universidade do Porto

Master of Science

CPT, Modified Gravity and the Baryon Asymmetry of the Universe

by Maria Pestana da Luz Pereira Ramos

The long-standing problem of the asymmetry between matter and antimatter in the Universe establishes a deep connection between Cosmology, Quantum Field Theory and Particle Physics. In this work, we derive a proof of the CPT Theorem, whose application justifies the existence and properties of antiparticles. Furthermore, we show that the identification of CPT with the strong reflection operation allows to formulate a framework where the Standard Model is extended to include Lorentz-violating terms, which can lead to CPT-violating observable effects. We also review the present status of observations explaining the observed baryon asymmetry, η, and the theoretical conditions for baryogenesis.

We study a class of models entailing violation of CPT symmetry in the early Universe, which allow the generation of η while baryon violating forces are still in equilibrium. This can occur due to space-time backgrounds which do not respect some of the assumptions for the validity of the CPT theorem. In particular, we derive the amount of η that can be produced by the classical motion of a scalar field in its potential, following Cohen and Kaplan's model for spontaneous baryogenesis. Another way to generate the observed asymmetry is to resort to modified gravity models, where the scalar field is replaced by the Ricci scalar: we argue that $f(R)$ gravity with a nonminimal geometry-matter coupling (NMC) naturally points toward the transition from spontaneous to gravitational baryogenesis.

As a generalization of current models of $f(R)$-baryogenesis, we proceed by examining the allowed region of variation for the mass scales and the exponents of both $f(R)$ and NMC functions, showing that tiny deviations from General Relativity are consistent with the observed baryon asymmetry and lead to temperatures compatible with Big Bang Nucleosynthesis.

UNIVERSIDADE DO PORTO

Resumo

Departamento de Física e Astronomia
Faculdade de Ciências da Universidade do Porto

Mestre de Ciência

CPT, Gravidade Modificada e a Assimetria Bariónica do Universo

por Maria Pestana da Luz Pereira Ramos

A existência da assimetria cosmológica entre matéria e antimatéria, um problema frequentemente discutido na literatura, estabelece uma conexão profunda entre a Cosmologia, a Teoria Quântica de Campo e a Física de Partículas. Neste trabalho, deriva-se uma prova do Teorema CPT, que justifica a existência e as propriedades das antipartículas. Em particular, identificando esta operação com a reflexão forte, é possível generalizar o Modelo Padrão, de modo a incluir termos que violam a simetria de Lorentz, alguns dos quais quebram também a invariância CPT, levando a efeitos possivelmente acessíveis à experiência. Para além disso, procede-se à revisão do estado corrente de observações que explicam a assimetria observacional, η, e das condições teóricas necessárias à bariogénese.

Estuda-se uma classe de modelos que admite uma violação da simetria CPT no Universo primitivo, permitindo a geração de η, enquanto as interações que violam o número bariónico estão ainda em equilíbrio térmico. Este cenário é possível devido à introdução de potenciais, interpretados como fundos cósmicos, que não respeitam alguma das hipóteses do Teorema CPT. Especificamente, avalia-se a quantidade η produzida devido ao movimento clássico de um campo escalar ao longo de um potencial, seguindo o modelo proposto por Cohen e Kaplan para o desenvolvimento de bariogénese espontânea. Uma forma alternativa de gerar a assimetria bariónica seria recorrer a modelos de gravidade modificada, substituindo o campo escalar pelo escalar de Ricci: deste modo, argumenta-se que as teorias $f(R)$ com um acoplamento não mínimo entre curvatura e matéria (NMC) apontam, naturalmente, para a transição de bariogénese espontânea para gravitacional.

Generalizando trabalhos prévios, que desenvolvem um mecanismo para a bariogénese no contexto das teorias $f(R)$, examina-se a região de variação das escalas de massa e dos expoentes de ambas as funções $f(R)$ e NMC, concluindo que pequenos desvios da Relatividade Geral são consistentes com a assimetria bariónica observada, sendo compatíveis com temperaturas características da Nucleossíntese Primordial.

Contents

List of Figures

List of Tables

List of Abbreviations

GR	**G**eneral **R**elativity
NMC	**N**on-**M**inimal **C**oupling
FRW	**F**riedmann **R**obertson-**W**alker
SM	**S**tandard **M**odel
BB	**B**ig **B**ang
BBN	**B**ig **B**ang **N**ucleosynthesis
CMB	**C**osmic **M**icrowave **B**ackground
C	**C**harge Conjugation transformation
P	**P**artity transformation
T	**T**ime Reversal transformation
SME	**S**tandard **M**odel **E**xtension
B	**B**aryon number
SR	**S**trong **R**eflection
HC	**H**ermitian **C**onjugation
WMAP	**W**ilkinson **M**icrowave **A**nisotropy **P**robe
JBD	**J**ordan-**B**rans-**D**icke

List of Symbols

η	Baryon-to-photon ratio
η_S	Baryon-to-entropy ratio
n_B	Net particle number
Γ	Decay width
T_F	Freeze-out temperature
T_D	Decoupling temperature
g	Degrees of freedom
g_*	Relativistic degrees of freedom
μ	Chemical potential
ρ	Energy density
p	Pressure
n	Particle Number
s	Entropy density
$T^{\mu\nu}$	Energy-momentum tensor
$g_{\mu\nu}$	Spacetime metric
R	Ricci Scalar
$\mathcal{L}_m$	Matter Lagrangian density

The physical constants $c, k_B, \hbar$ are usually set to 1.

The metric signature is (+ - - -).

Chapter 1

Introduction

One of the major mysteries of our Universe is that we do not observe the same number of particles and antiparticles. In fact, the observed Universe is entirely made of matter, except for a tiny percentage of antiparticles detected in the cosmic ray flux.

Starting from a neutral and symmetric Universe, we cannot explain this asymmetry. This is based on the fact that every well-defined relativistic quantum field theory that we have constructed respects CPT invariance, which — in turn — identifies every particle state with an antiparticle one.

Baryogenesis is a complicated problem, with many open questions, and the subject of intensive research. It connects different areas of physics beyond elementary particle theory: on one hand, it is constrained by cosmological parameters and the sequential phases of expansion; on the other hand, quantum field theory is an obvious prerequisite to define the state of an antiparticle and to study effective theories. Moreover, this problem requires various concepts of thermodynamics and, in particular, a thermodynamic formulation applied to the early Universe, an important topic that we will discuss in this work.

Sakharov (1967) proposed the three necessary conditions that should be required by all baryogenesis models. Since known mechanisms are not able to generate enough matter-antimatter asymmetry, extensions of the Standard Model are required. We could have focused on reviewing the specific features of each model and its particular problems — a nice review on these topics can be found, for example, in Ref. [1] — ; instead, we chose to explore an alternative to the standard conditions, that allows us to discuss the CPT invariance in a cosmological context.

In fact, faced with the consequences of the CPT theorem, a natural way to address the open question of how an expanding Universe could develop an excess of baryons over antibaryons is to argue that CPT might not have been already established in the primary stages of evolution. This allows for a different particle-antiparticle interaction with the cosmological background, while ensuring the standard evolution that leads to a CPT-invariant ground state.

Moreover, bringing modified gravity models to the discussion, we obtain modified field equations that can enlarge the baryonic asymmetry to the observable amount, while enabling us to interpret thermodynamically the new conservation laws. In particular, entropy considerations might be altered, which can significantly affect the cosmological parameter that characterizes this asymmetry.

The present mechanism for baryogenesis is an interesting framework to connect various sub-fields of physics and to extend basic concepts of microphysics to the cosmological system. Aside from being a rich problem to learn (not only the theory, but how this is compatible with observations and particle experiments), this model allows us to complete and generalize previous work, hence contributing to a deeper comprehension of gravitational baryogenesis.

This thesis is divided into four major parts:

1. The **research topics**, where Chapter 1 and Chapter 2 are included, regarding CPT invariance and the thermal history of the Universe, where we review the status of observations leading to the numerical value for the asymmetry parameter and the cosmological reasons for this to be the baryon-to-entropy ratio;

2. The discussion of two **existing models that incorporate CPT violation** in an expanding Universe, corresponding to spontaneous and gravitational baryogenesis. These models are studied in Chapter 4 and Chapter 6, respectively;

3. The presentation, in Chapter 5, of the ***f(R)*-model** with the inclusion of a **nonminimal geometry-matter coupling** (NMC) and the attractive features of the latter;

4. The **new results**, which are discussed also in Chapter 6, aiming to develop a model for the generation of the baryon asymmetry, in the framework of modified gravity theories. This generalizes previous works, where only the $f(R)$-function was included, and explores the impact of the NMC in the mechanism leading to gravitational baryogenesis. This chapter is based upon the work reported in Ref. [2].

At the beginning of each chapter, we describe — with more detail — the sequential line of thought along this thesis.

Finally, in Chapter 7, we present our conclusions and prospects of future work.

Chapter 2

CPT Invariance

As the motivation for our work is a dynamical violation of CPT, we begin by exploring this invariance.

We derive a simple proof of the CPT theorem. This is a very interesting result and maybe the most important one in what concerns the predictions of antiparticles. It states that a wide class of quantized field theories is invariant with respect to the product of time reserval (**T**), charge conjugation (**C**) and parity (**P**).

The aim of this proof is not to focus solely on mathematical techniques, but to discuss the physics behind it. We will consider three types of fields - spin 0, spin 1/2 and spin 1. Given the generality of the theorem, it is important to know where each assumption enters the proof to gain intuition on how it could possibly be broken.

Later on, we justify several properties of particles/antiparticles that we will use in the course of the work and which are consequence of the CPT Theorem. Finally, we address some questions regarding CPT-violation phenomenology and experimental constraints.

2.1 Field Operators

Let us start by listing the equations for the various types of fields we will discuss.

Spin-0 fields are described by the Klein-Gordon equation:

$$(\Box + m^2)\phi(x) = 0 \ , \quad (\Box + m^2)\phi^*(x) = 0 \ ; \tag{2.1.1}$$

expanding the fields in terms of the creation and annihilation operators, we can derive the free field commutation relations:

$$[\phi(x), \phi^*(y)] = i\Delta(x-y) \ , \quad [\phi(x), \phi(y)] = [\phi^*(x), \phi^*(y)] = 0 \ . \tag{2.1.2}$$

Spin-$\frac{1}{2}$ fields are described by the Dirac equation:

$$(i\nabla\!\!\!/ - m)\psi(x) = 0 \ , \quad \overline{\psi}(x)(-i\overleftarrow{\nabla\!\!\!/} - m) = 0 \ ; \tag{2.1.3}$$

similarly, the anticommutation relations read

$$\{\psi_\alpha(x), \overline{\psi}_\beta(y)\} = i(i\nabla\!\!\!/ + m)_{\alpha\beta}\Delta(x-y) \ , \quad \{\psi_\alpha(x), \psi_\beta(y)\} = \{\overline{\psi}_\alpha(x), \overline{\psi}_\beta(y)\} = 0 \ , \tag{2.1.4}$$

where

$$\overline{\psi} = \psi^\dagger\gamma^0 \ , \tag{2.1.5}$$

$$\gamma^\mu\gamma^\nu + \gamma^\nu\gamma^\mu = 2g^{\mu\nu} \ , \tag{2.1.6}$$

with γ denoting the 4×4 Dirac's matrices. Since we are using the metric (+,-,-,-), it follows that

$$\gamma^{0\dagger} = \gamma^0 \ , \quad \gamma^{i\dagger} = -\gamma^i \ (i = 1,2,3) \ . \tag{2.1.7}$$

Spin-1 fields, in each component, also satisfy the Klein-Gordon equation:

$$(\Box + m^2)\phi_\mu(x) = 0 \ , \quad (\Box + m^2)\phi^*_\mu(x) = 0 \ ; \tag{2.1.8}$$

whereas the commutation relations read

$$[\phi_\mu(x), \phi^*_\nu(y)] = -i\left(g_{\mu\nu} + \frac{1}{m^2}\partial_\mu\partial_\nu\right)\Delta(x-y) \ , \quad [\phi_\mu(x), \phi_\nu(y)] = [\phi^*_\mu(x), \phi^*_\nu(y)] = 0 \ . \tag{2.1.9}$$

For sake of simplicity, we used the same symbol m for all the masses appearing not only in the field equations, but also in the commutator function.

The Δ function is defined in appendix B and is calculated explicitly for the case of a spin-0 field. The other cases can be found, for example, in Greiner's *Field Quantization* [3]. The derivation of the expression for $\Delta(x)$ will allow us to discuss its symmetry properties and consequently the invariance of the transformed laws, subjected to the action of some symmetry operator.

2.2 Discrete Symmetries

If we want to understand the CPT Theorem, first we need to define each of these operations, which must leave the field equations and commutation relations invariant.

2.2.1 Parity

To discuss parity, let us start by taking into account the existence of the improper Lorentz transformation of space reflection:

$$\boldsymbol{x'} = -\boldsymbol{x} \ , \quad t' = t \ . \tag{2.2.10}$$

The corresponding transformation matrix is

$$\Lambda^{\mu\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} = g^{\mu\nu} \ . \tag{2.2.11}$$

In order to establish Lorentz covariance of the Dirac equation, we introduce a matrix such that $\psi'(x') = \psi'(\Lambda x) = S(\Lambda)\psi(x)$; then, after a Lorentz transformation, (2.1.3) remains invariant provided an S can be found which has the property $S(\Lambda)\gamma^\mu S^{-1}(\Lambda)\Lambda_\mu^\nu = \gamma^\nu$. Denoting $S = P$ for the coordinate reflection, this condition becomes

$$P\gamma^\mu P^{-1} = \gamma_\mu \ , \tag{2.2.12}$$

which is satisfied by

$$P = e^{i\alpha}\gamma^0 \ . \tag{2.2.13}$$

In spite of the arbitrariness in the choice of the phase factor, we have to be coherent in the choice of the intrinsic parity of particles (for example, if we consider the exchange of pions between nucleons, $p + n \to \pi$, which conserves parity, we need to guarantee that the intrinsic parity of the proton times the intrinsic parity of the neutron equals the intrinsic parity of the pion). Generally, we define:

$$\boldsymbol{P}\psi(\boldsymbol{r},t)\boldsymbol{P}^{-1} = \alpha''_P\gamma^0\psi(-\boldsymbol{r},t) \ . \tag{2.2.14}$$

Under this parity transformation, it is easy to show that the transformed field, $\psi'(x) = \alpha''_P\gamma^0\psi(-\boldsymbol{r},t)$ also satisfies the Dirac equation.

For the free Klein-Gordon theory, the condition

$$\boldsymbol{P}\phi(\boldsymbol{r},t)\boldsymbol{P}^{-1} = \alpha_P\phi(-\boldsymbol{r},t) \ , \tag{2.2.15}$$

satisfies

$$\boldsymbol{P}\mathcal{L}(\boldsymbol{r},t)\boldsymbol{P}^{-1} = \mathcal{L}(-\boldsymbol{r},t) \ . \tag{2.2.16}$$

Moreover, if we include an electromagnetic current, that is, if we add to the Lagrangian density a term representing the interaction, generally electromagnetic, of the quantum system with an external field,

$$\mathcal{L} \to \mathcal{L} - j_\mu A^\mu \ , \tag{2.2.17}$$

we can verify that (2.2.15) ensures

$$\boldsymbol{P}j_\mu(\boldsymbol{r},t)\boldsymbol{P}^{-1} = j^\mu(-\boldsymbol{r},t) \ , \tag{2.2.18}$$

so the equations of motion remain unchanged. This is true because, working with the wave equation (2.1.1), we were able to identify the conserved current with $j^\mu = i(\phi^*\nabla^\mu\phi - \phi\nabla^\mu\phi^*)$. This parity transformation also leaves the commutation relations invariant.

Finally, in the case of the spin-1 field, the spatial components transform as a vector under spatial reflection, while the time component remains unaffected; thus

$$\boldsymbol{P}\phi_i(\boldsymbol{r},t)\boldsymbol{P}^{-1} = -\alpha'_P\phi_i(-\boldsymbol{r},t) \ , \tag{2.2.19}$$

$$\boldsymbol{P}\phi_0(\boldsymbol{r},t)\boldsymbol{P}^{-1} = \alpha'_P\phi_0(-\boldsymbol{r},t) \ . \tag{2.2.20}$$

2.2.2 Charge Conjugation

Now we focus on the charge conjugation operation, a symmetry that emerges from the fact that to each particle, there is an antiparticle. In particular, it sustains the idea that the existence of electrons implies the existence of positrons, which was a crucial aspect to understand the hole theory that emerged from the negative solutions of the Dirac equation. To avoid these unphysical solutions, Dirac proposed that the vacuum of the theory corresponds to a configuration where all negative energy states are occupied, forming the so-called "Dirac sea"; positive energy electrons are then forbidden to fall into this fully occupied state by Pauli's Exclusion Principle. On the contrary, a photon could excite an electron from a negative energy state, leaving a hole in the vacuum. The physical interpretation is that this hole appearing in the absence of an energy $-E$ $(E > 0)$ and a charge equal to e $(e < 0)$ is equivalent to the presence of a positron with positive energy $+E$ and charge $-e$. The modern version of this picture is understood in terms of the Feynman-Stückelberg interpretation. The point we wish to underline is that there is a one-to-one (experimentally established) correspondence between the negative solutions of the Dirac equation

$$(i\not{\nabla} - e\not{A} - m)\psi = 0 \ , \tag{2.2.21}$$

and the positron wave function, ψ_c. Following this interpretation, positrons appear as positively charged electrons, so ψ_c will be a positive-energy solution of the equation

$$(i\not{\nabla} + e\not{A} - m)\psi_c = 0 \ . \tag{2.2.22}$$

There should be a transformation which, starting with equation (2.2.21), would lead us to the existence of an antiparticle satisfying equation (2.2.22). Taking the complex conjugate of the first, we obtain

$$[(i\partial_\mu + eA_\mu)\gamma^{\mu *} + m]\psi^* = 0 \ , \tag{2.2.23}$$

with $A_\mu = A_\mu^*$. If we can now find a nonsingular matrix, that we denote $C\gamma^0$, with the algebra

$$(C\gamma^0)\gamma^{\mu *}(C\gamma^0)^{-1} = -\gamma^\mu \ , \tag{2.2.24}$$

we will find the desired form

$$(i\not{\nabla} + e\not{A} - m)(C\gamma^0\psi^*) = 0 \ , \tag{2.2.25}$$

with

$$C\gamma^0\psi^* = C\overline{\psi}^T = \psi_c \tag{2.2.26}$$

the positron wave function. We may verify that such a matrix C indeed exists by explicit construction. In our representation (C.0.1 and C.0.2), $\gamma^0\gamma^{\mu *}\gamma^0 = \gamma^{\mu T}$, so that condition (2.2.24) becomes $C\gamma^{\mu T}C^{-1} = -\gamma^\mu$, or

$$C^{-1}\gamma^\mu C = -\gamma^{\mu T} \ . \tag{2.2.27}$$

In this representation, C must commute with γ^1 and γ^3 and anticommute with γ^0 and γ^2; this is easily satisfied for

$$C = i\gamma^2\gamma^0 = -C^{-1} = -C^\dagger = -C^T \ . \tag{2.2.28}$$

It is enough to be able to construct a matrix C in any given representation; applying a unitary transformation to any other one will give a matrix appropriate to the new representation. The definite operation has the form

$$\boldsymbol{C}\psi(\boldsymbol{r},t)\boldsymbol{C}^{-1} = \alpha''_C C\overline{\psi}^T \ , \tag{2.2.29}$$

$$\boldsymbol{C}\overline{\psi}(\boldsymbol{r},t)\boldsymbol{C}^{-1} = -\alpha''^*_C \psi^T C^\dagger \ . \tag{2.2.30}$$

There is a phase arbitrariness in our definitions, just like in the case of a parity transformation.

The operator (2.2.26) explicitly constructs the wave function of a positron. However, we would like to develop from it an invariance operation for the Dirac equation. This is possible if the charge conjugation transformation additionally changes the sign of the electromagnetic field. Then, the sequence of instructions

1. take the complex conjugate ;
2. multiply by $C\gamma^0$;
3. replace A_μ by $-A_\mu$;

defines a *formal symmetry operation of the Dirac theory*. This means that, for each physically realizable state containing an electron in a potential $A_\mu(x)$, there corresponds a physical realizable state of a positron in a potential $-A_\mu(x)$. Although we already knew the equivalence of these two dynamics from classical considerations, we are led to a new and surprising result: if there exist electrons of mass m and charge e, there must also exist positrons of mass m and charge $-e$[1].

When applying charge conjugation to the Lagrangian, in order to check its invariance, one encounters the difficulty that products of operators are transformed into those of the Hermitian adjoints:

$$\boldsymbol{C}\overline{\psi}(x)\gamma^\mu\psi(x)\boldsymbol{C}^{-1} = -\psi_\alpha(x)C^{-1}_{\alpha\beta}\gamma^\mu_{\beta\lambda}C_{\lambda\tau}\overline{\psi}_\tau(x) = \psi_\alpha(x)\gamma^\mu_{\tau\alpha}\overline{\psi}_\tau(x) \ . \tag{2.2.31}$$

In order to make the transformed Lagrangian comparable with the original one, we need to assume from the beginning that a process of proper symmetrization (or equivalently normal ordering) has been

[1]This conclusion is a consequence of the identification of (2.2.26) with the positron field. Later on, the state of an antiparticle is generalized to be $|\psi'\rangle \equiv CPT\,|\psi\rangle$, as charge conjugation can be violated and the consequences of the CPT theorem hold still.

applied to it, so that all quantities are symmetrized in their boson fields and antisymmetrized in the fermion fields. Identifying the electromagnetic current with

$$j^\mu(x) = \frac{1}{2}\left[\overline{\psi}(x), \gamma^\mu \psi(x)\right] \ , \tag{2.2.32}$$

it follows directly from (2.2.31) that $\boldsymbol{C} j^\mu(x) \boldsymbol{C}^{-1} = -j^\mu(x)$, and therefore this quantity is odd under charge conjugation transformation, as it should: reversing the role of particle and antiparticle reverses the electromagnetic field; then, the electromagnetic current must also gain a minus sign, in order to preserve the invariance of the $(A_\mu j^\mu)$ interaction.

We saved for last the transformation of boson fields. In terms of the Lagrangian (2.2.17), the previous discussion leads to the requirement, for a theory which is charge conjugation invariant, that there exists a unitary operator such that

$$\boldsymbol{C}\mathcal{L}(x)\boldsymbol{C}^{-1} = \mathcal{L}(x) \ , \tag{2.2.33}$$

and

$$\boldsymbol{C} j_\mu(x) \boldsymbol{C}^{-1} = -j_\mu(x) \ . \tag{2.2.34}$$

Since $\mathcal{L}(x) \to \mathcal{L}(x)$ and $j_\mu(x) \to -j_\mu(x)$ under the transformation $\phi \leftrightarrow \phi^*$, we search for a $\boldsymbol{C}$ which has the property

$$\boldsymbol{C}\phi(\boldsymbol{r}, t)\boldsymbol{C}^{-1} = \alpha_C \phi^*(\boldsymbol{r}, t) \ . \tag{2.2.35}$$

Again, the generalization for spin-1 fields follows the same principles:

$$\boldsymbol{C}\phi_\mu(\boldsymbol{r}, t)\boldsymbol{C}^{-1} = \alpha'_C \phi^*_\mu(\boldsymbol{r}, t) \ . \tag{2.2.36}$$

Note, for completeness, that $\boldsymbol{C}$ must also change electrically neutral particles, described by non-hermitian fields, such as the neutron (n), to their antiparticles $(\bar{n})$, in order that the conservation laws of strangeness, nucleon number and isospin are invariant under $\boldsymbol{C}$. For hermitian fields, like those describing photons, the commutation relations vanish and the particle is not distinguishable from the antiparticle; under $\boldsymbol{C}$, the hermitian field can at most change by a factor -1. As discussed, in the case of the electromagnetic field, A_μ must transform according to:

$$\boldsymbol{C} A^\mu(x) \boldsymbol{C}^{-1} = -A^\mu(x) \ , \tag{2.2.37}$$

in order to leave the Lagrangian density invariant.

Conventionally, $\boldsymbol{C}\,|0\rangle = +\,|0\rangle$ is postulated, as well as $\boldsymbol{P}\,|0\rangle = +\,|0\rangle$, for the free field vacuum $|0\rangle$; that means the vacuum is an even eigenstate. Also, for an $n-$photon state, the eigenvalue of $\boldsymbol{C}$ according to (2.2.37) is $(-1)^n$. This is important to understand the consistency of the phase in our definitions. Consider, for example, the observation of the decay $\pi^0 \to 2\gamma$. If charge conjugation invariance holds in strong and electromagnetic interactions, then π^0 must be even under $\boldsymbol{C}$ if it evolves to the state of $\gamma\gamma$, which — by equation (2.2.37) — is even.

2.2.3 Time Reversal

If we transform the electron wave function under time reversal, we will obtain the original electron running backwards in time. This state will be physically realizable, provided that the transformed wave function ψ' also satisfies the Dirac equation. Hence, we need to find an operator $\boldsymbol{T}$ which transforms physical states evolving in time t to states as would be viewed on backwards, with $t' = -t$.

To begin our discussion, consider the Heisenberg equation

$$[\mathcal{H}, \psi_m(\boldsymbol{r}, t)] = -i\frac{\partial \psi_m(\boldsymbol{r}, t)}{\partial t} \ . \tag{2.2.38}$$

If we seek a unitary operator $\boldsymbol{U}$ which leaves the action invariant and which transforms $\psi_m(\boldsymbol{r}, t)$ to $W_{mn}\psi_n(\boldsymbol{r}, -t) = \boldsymbol{U}\psi_m(\boldsymbol{r}, t)\boldsymbol{U}^{-1}$, we obtain

$$[\boldsymbol{U}\mathcal{H}\boldsymbol{U}^{-1}, \psi_n(\boldsymbol{r}, t')] = i\frac{\partial \psi_n(\boldsymbol{r}, t')}{\partial t'} \tag{2.2.39}$$

In order to restore equation (2.2.38), we would need $\boldsymbol{U}$ to transform $\mathcal{H}$ to $-\mathcal{H}$. This is unacceptable in physics, because it would mean that, after the transformation, the eigenvalues of $\mathcal{H}$ would be negative relative to the vacuum state (the energy would be unbounded from below). To avoid this situation, $\boldsymbol{T}$ is considered an antiunitary operator, that takes the complex conjugate of all c-numbers involved,

$$\boldsymbol{T}i\boldsymbol{T}^{-1} = -i \ . \tag{2.2.40}$$

With this choice, (2.2.38) will be invariant under $\boldsymbol{T}$. In terms of the Lagrangian density (2.2.17), the theory will be time-reversal invariant if

$$\boldsymbol{T}\mathcal{L}(\boldsymbol{r}, t)\boldsymbol{T}^{-1} = \mathcal{L}(\boldsymbol{r}, -t) \ , \tag{2.2.41}$$

and

$$\boldsymbol{T}j_\mu(\boldsymbol{r}, t)\boldsymbol{T}^{-1} = j^\mu(\boldsymbol{r}, -t) \ , \tag{2.2.42}$$

meaning the electromagnetic currents are reversed while the charges are unchanged. Also, to guarantee the invariance of the Lagrangian, A^μ should transform like:

$$A_\mu(\boldsymbol{r}, t) \to A^\mu(\boldsymbol{r}, -t) \ . \tag{2.2.43}$$

In particular, $\overrightarrow{A}'(t') = -\overrightarrow{A}(t)$. Also, $\overrightarrow{\nabla}' = \overrightarrow{\nabla}$ and $\overrightarrow{x}' = \overrightarrow{x}$. It is clear by now that the time-reversal transformation changes i to $-i$; hence, $\boldsymbol{T}$ may be equivalent to taking the complex conjugate and then multiplying by a 4×4 constant matrix T:

$$\psi'(t') = T\psi^*(t) \ . \tag{2.2.44}$$

To construct the desired transformation, we consider the Dirac equation in the presence of an external electromagnetic field[2],

$$[(i\nabla_\mu - eA_\mu)\gamma^\mu - m]\,\psi = 0 \ . \tag{2.2.45}$$

[2]The coupling with the electromagnetic field is simply introduced by means of the substitution $p^\mu \to p^\mu - eA^\mu$, where $p^\mu = i\nabla^\mu = i\frac{\partial}{\partial x_\mu}$ and $x^\mu = (t, \overrightarrow{x})$.

Doing the operations we have just dictated, we get[3]

$$\left[i(T\gamma^{0*}T^{-1})\frac{\partial}{\partial t'} - i(T\gamma^{i*}T^{-1})\frac{\partial}{\partial x'^{i}} - e(T\gamma^{0*}T^{-1})A_0' + e(T\gamma^{i*}T^{-1})A_i' - m\right]\psi'(t') = 0 \ . \tag{2.2.46}$$

For ψ' to satisfy the original Dirac equation, $T\gamma^{0*}T^{-1} = \gamma^0$ while $T\gamma^{i*}T^{-1} = -\gamma^i$. In our representation (C.0.1 and C.0.2), since $\gamma^{0*} = \gamma^0$, $\gamma^{1*} = \gamma^1$, $\gamma^{2*} = -\gamma^2$ and $\gamma^{3*} = \gamma^3$, this means that the T-matrix must commute with γ^0 and γ^2 and anticommute with γ^1 and γ^3; a suitable choice is

$$T = i\gamma^1\gamma^3 = T^\dagger = T^{-1} = -T^* \ , \tag{2.2.47}$$

where the phase factor is again arbitrary. An equivalent statement of this condition is the requirement that

$$T\gamma^\mu T^{-1} = \gamma^{\mu T} \ , \tag{2.2.48}$$

since $\gamma^{0T} = \gamma^0$, $\gamma^{1T} = -\gamma^1$, $\gamma^{2T} = \gamma^2 = -\gamma^{2*}$ and $\gamma^{3T} = -\gamma^3$, in the usual representation.

Although the property (2.2.44) makes sense in one-particle theory, upon the fact that $\boldsymbol{T}$ takes a $c-$number to its complex conjugate, the transformation $\psi \to T\psi^\dagger$ is unacceptable in field theory, because this would transform an electron at rest into a positron state. So, for the Dirac theory, we must redefine the operator $\boldsymbol{T}$ such that

$$\boldsymbol{T}\psi(\boldsymbol{r},t)\boldsymbol{T}^{-1} = \alpha_T'' T\psi(\boldsymbol{r},-t) \ , \tag{2.2.49}$$

with T being the same matrix we found before. It is easy to see that this matrix satisfies both equations (2.2.41) and (2.2.42); for example, to prove the invariance of the Dirac Lagrangian, note that — after the time reversal transformation — we get:

$$\begin{aligned}
\boldsymbol{T}\mathcal{L}(\boldsymbol{r},t)\boldsymbol{T}^{-1} &= \boldsymbol{T}\overline{\psi}(i\nabla\!\!\!/ - m)\psi\boldsymbol{T}^{-1} \\
&= \overline{\psi}'T\left(-i\gamma^{*0}\frac{\partial}{\partial(-t')} - i\boldsymbol{\gamma^*}\cdot\boldsymbol{\nabla}' - m\right)T^{-1}\psi' \\
&= \overline{\psi}'\left(i(T\gamma^{*0}T^{-1})\frac{\partial}{\partial t'} - i(T\boldsymbol{\gamma}^*T^{-1})\cdot\boldsymbol{\nabla}' - m\right)\psi' \\
&= \overline{\psi}'\left(i\gamma^0\frac{\partial}{\partial t'} + i\boldsymbol{\gamma}\cdot\boldsymbol{\nabla}' - m\right)\psi' \\
&= \mathcal{L}(\boldsymbol{r'},t') = \mathcal{L}(\boldsymbol{r},-t) \ ,
\end{aligned} \tag{2.2.50}$$

so equation (2.2.41) is satisfied.

For the Klein-Gordon field, it is easy to see that the invariance laws (2.2.41) and (2.2.42) are satisfied provided

$$\boldsymbol{T}\phi(\boldsymbol{r},t)\boldsymbol{T}^{-1} = \alpha_T\phi(\boldsymbol{r},-t) \ . \tag{2.2.51}$$

The transformation of a spin-1 field is also rather intuitive:

$$\boldsymbol{T}\phi_0(\boldsymbol{r},t)\boldsymbol{T}^{-1} = -\alpha_T'\phi_0(\boldsymbol{r},-t) \ ; \tag{2.2.52}$$

[3] Be aware that it is the operation under time reversal that changes $i \to -i$, not the multiplication by the T matrix.

$$\boldsymbol{T}\phi_i(\boldsymbol{r},t)\boldsymbol{T}^{-1} = \alpha'_T\phi_i(\boldsymbol{r},-t) \ . \tag{2.2.53}$$

Let us now analyze the behavior of the various fields. The equations of motion of spin-zero and spin-one fields are easily seen to be preserved. But recalling the symmetry properties of the Δ function (see appendix B), we notice it changes sign under the substitution $t \to -t$. Again, this only means that $\boldsymbol{T}$ cannot be a linear operator. Since the r.h.s of the commutation relations are purely imaginary, it suffices to define $\boldsymbol{T}$ as an antilinear operator, as argued when discussing the Heisenberg equation for the evolution of fields.

Another important property to notice is that the modulus of the c-numbers, α_U, α'_U and α''_U with $U = C, P, T$, must be equal to one, in order to conserve the commutation relations.

Note also that combining equations (2.2.27) and (2.2.48), we get the following result

$$(CT)\gamma^\mu(CT)^{-1} = -\gamma^\mu \ , \tag{2.2.54}$$

which shows that, since $\{\gamma^\mu, \gamma^5\} = 0$, we may put

$$CT = \gamma^5 \ . \tag{2.2.55}$$

2.3 Proof of the CPT Theorem

In 1957, Lüders [4] proved that a wide class of field theories invariant under the proper Lorentz group (the field operator expansion and the commutation relations are built to be Lorentz invariant) is also invariant under the product **CPT**. This theorem does not state that this product is identical to the unit operator, but that the c-numbers $\alpha_T, \alpha_C, \alpha_P$, etc., can be chosen in a way that one has invariance. Lüders proof, which we will follow up, consists on explicitly constructing an operation under which the theory is invariant and then showing that this is equivalent to the product **CPT**. This operation is done in two steps: the strong reflection and the subsequent hermitian conjugation.

2.3.1 Strong Reflection

Working out permutations of **CPT** on specific examples, we can convince ourselves that it is actually quite difficult to construct a Hamiltonian that violates the product of **C**, **P** and **T** taken in any order (with suitable choices of the phases)[4]. This fact induces us to think that there is something fundamental about the product of these three symmetries. Then we are naturally led to ask what are the minimal assumptions on a field theory of elementary particles, in order that this product is a *good* operation that leaves it invariant. If we think about it, as a theoretical physicist that is faced with this symmetry hint, we would start by assuming invariance under proper orthochronous Lorentz transformations, since all objects

[4]This idea is explored in the fifth chapter of Sakurai's book [5], for example.

in field theory (the operators itself, the field equations, the commutation relations and propagators, the volume elements on integrals, etc.) are constructed to preserve the relativistic laws. As we go along, it will become necessary to make a few more assumptions, like the usual spin-statistics connection.

Before considering the product **CPT**, let us enquire if there is any transformation such that the transformation property depends only on the oddness or evenness of the rank (number of uncontracted indexes) of a tensor density or some other quantity. That is, if there exists, for a tensor with n uncontracted Lorentz indexes, any transformation that takes

$$\mathcal{O}_{\mu\nu\ldots\lambda\sigma}(x) \to (-1)^n \mathcal{O}_{\mu\nu\ldots\lambda\sigma}(-x) \ . \tag{2.3.56}$$

If we can find a set of operations that lead to (2.3.56), then anything we can write down for an interaction density will be accordingly invariant under that set of operations. This would be true for both the Lagrangian density, which is a *true* scalar, and for the Hamiltonian density, which is the 0-0 component of the energy-momentum tensor, a symmetric tensor of rank 2. Note also that, for the 4-current, transformation (2.3.56) means $j_\mu(x) \to -j_\mu(-x)$, which cannot be brought by either **C**, **P** or **T** alone. We now want to show that the desired set of operations called "strong reflection" exists and is closely related to the product **CPT** (even though this is not the whole story).

We only assume Lorentz invariance — not reflection invariance under parity nor time reversal — so that the rank of a tensor is defined exclusively from its behavior under the proper orthochronous Lorentz transformations[5], $L_+^\uparrow$:

$$\mathcal{O}_{\mu\nu\ldots\lambda\sigma} \to \Lambda^{\mu'}{}_{\mu}\Lambda^{\nu'}{}_{\nu}\ldots\Lambda^{\lambda'}{}_{\lambda}\Lambda^{\sigma'}{}_{\sigma}\mathcal{O}_{\mu'\nu'\ldots\lambda'\sigma'} \ , \tag{2.3.57}$$

where $\Lambda^{\mu'}{}_{\mu}$ stands for the matrix element associated with the Lorentz transformation $x_\mu \to \Lambda^{\mu'}{}_{\mu}x_{\mu'}$.

Now let us wonder what is the physical meaning of the transformation (2.3.56). If 4-dimensional space were Euclidean, there would be only two classes of Lorentz transformations[6] , one with $\det\Lambda^\mu{}_\nu = 1$ and the other one with $\det\Lambda^\mu{}_\nu = -1$. In particular, the transformation

$$x \to -x \ , \quad \Lambda^\mu{}_\nu = \begin{pmatrix} -1 & & & 0 \\ & -1 & & \\ & & -1 & \\ 0 & & & -1 \end{pmatrix} \tag{2.3.58}$$

is of the first type and can be generated by continuous rotations in 4-space: simply rotate in the $3-4$ plane about the $1-2$ plane by π and then rotate in the $1-2$ plane about the $3-4$ plane by π (here,

[5]The proper orthochronous Lorentz group is defined by the condition $L_+^\uparrow = \{\Lambda \in L : \det\Lambda = +1, \Lambda_0^0 \geq 1\}$. This subgroup is the one (out of four) component of the Lorentz group that contains all Lorentz transformations that can be connected to the identity by a continuous curve lying in the group. It is generated by ordinary spatial rotations and Lorentz boosts. It can be shown that if M denotes a general matrix (with ordinary group operations) with $\det(M) = 1$, then $\Lambda(M) \in L_+^\uparrow$.

[6]For 4-dimensional Euclidean space, the metric becomes $g_{\mu\nu} = \delta_{\mu\nu}$; hence, the condition on the Lorentz matrix, $\Lambda^T g\Lambda = g$, reduces to $\Lambda^\mu{}_0\Lambda^\mu{}_0 = 1$, which implies $(\Lambda^0{}_0)^2 = 1 - (\Lambda^i{}_0)^2 \leq 1$ so that, for ordinary Lorentz transformations ($\Lambda^0{}_0, \Lambda^i{}_0$ reals), the usual two possible conditions on $\Lambda^0{}_0$ reduce to one: $-1 < (\Lambda^0{}_0) < 1$.

the 4 direction corresponds to the direction of time, that now behaves as any other spatial component). States with angular momentum J would transform as:

$$\begin{aligned} |\psi\rangle &\to S\,|\psi\rangle \ , \\ S &= e^{iJ_{12}\pi}e^{iJ_{34}\pi} \ , \end{aligned} \tag{2.3.59}$$

where J_{ij} are the generators of the rotation group. Recall that a general Lorentz transformation has the form

$$U_\omega = e^{i\frac{\omega^{\mu\nu}}{2}J_{\mu\nu}} \ , \tag{2.3.60}$$

where $\omega^{\mu\nu}$ are six independent parameters (three for boosts along each spatial direction and three for a generic rotation). The generators $J_{\mu\nu}$ include not only differential operators acting on the coordinate functions (the orbital part), but also the operators acting on the spinors, so they are generally written as

$$J_{\mu\nu} = i(x_\nu\partial_\mu - x_\mu\partial_\nu) + \frac{\sigma_{\mu\nu}}{2} \ . \tag{2.3.61}$$

We know that a scalar function is invariant under rotations and that a vector transforms as $\phi_i'(\boldsymbol{x}) = R_i^j\phi_j(\boldsymbol{R}^{-1}\boldsymbol{x})$; analogously, a solution of Dirac equation should transform as $\psi_a'(\boldsymbol{x}) = S_a^b\psi_b(\boldsymbol{\Lambda}^{-1}x)$, with S satisfying the condition

$$S^{-1}\gamma^\mu S = \Lambda^\mu_\nu\gamma^\nu \ , \tag{2.3.62}$$

to guarantee the covariance of the Dirac equation. Expanding (2.3.60) around the identity and using the fact that $\omega^{\mu\nu} = -\omega^{\nu\mu}$, yields

$$S(\omega) = 1 + \frac{i}{4}\omega^{\mu\nu}\sigma_{\mu\nu} + ... \tag{2.3.63}$$

and we can identify $\sigma_{\mu\nu}/2$ with the generators of the Lorentz group acting on the Dirac space. If we now use this result on the condition (2.3.62), we get a relation between these generators and the gamma matrices:

$$\sigma^{\mu\nu} = \frac{i}{2}[\gamma^\mu, \gamma^\nu] \ . \tag{2.3.64}$$

Returning to our discussion, under the transformation (2.3.58), the field operators transform as

$$\phi(x) \to \phi(-x) \ ; \tag{2.3.65}$$

$$\phi_\mu(x) \to -\phi_\mu(-x) \ ; \tag{2.3.66}$$

$$\psi(x) \to e^{i\frac{\pi}{2}\sigma_{12}}e^{i\frac{\pi}{2}\sigma_{34}}\psi(-x) \ . \tag{2.3.67}$$

The σ_{ij} operators in the exponent of equation (2.3.67) are found to be (see complement C):

$$\sigma^{12} = \begin{pmatrix} \sigma^3 & 0 \\ 0 & \sigma^3 \end{pmatrix} \ , \quad \sigma^{34} = -i\begin{pmatrix} 0 & \sigma^3 \\ \sigma^3 & 0 \end{pmatrix} \ . \tag{2.3.68}$$

Neglecting the unimportant phase factor in σ_{34} and since $\sigma_3^2 = 1$, we can expand the exponentials in terms of sine and cosine functions:

$$\psi(x) \to \left(\cos\frac{\pi}{2} + i\sigma_{12}\sin\frac{\pi}{2}\right)\left(\cos\frac{\pi}{2} + i\sigma_{34}\sin\frac{\pi}{2}\right)\psi(-x) = \gamma_5\psi(-x) \ . \tag{2.3.69}$$

In reality, however, the space is not Euclidean. The transformation $x \to -x$ belongs to $L_+^\downarrow$, because Λ_0^0 is negative and cannot be brought about continuously from the identity doing infinitesimal transformations. Even so, this gives a hint why the transformation $\psi \to \gamma_5\psi(-x)$ (up to a phase) has something to do with $\mathcal{O}_{\mu\nu\ldots\lambda\sigma} \to (-1)^n\mathcal{O}_{\mu\nu\ldots\lambda\sigma}$. It turns out the situation is a little less simple, and this operation has to be made in conjunction with another one. At this point, we define (as Lüders) the behaviour of the field operators under strong reflection as follows:

$$\phi(\boldsymbol{r},t) \to \phi(-\boldsymbol{r},-t), \quad \phi^*(\boldsymbol{r},t) \to \phi^*(-\boldsymbol{r},-t) \; ; \tag{2.3.70}$$

$$\phi_\mu(\boldsymbol{r},t) \to -\phi_\mu(-\boldsymbol{r},-t), \quad \phi_\mu^*(\boldsymbol{r},t) \to -\phi_\mu^*(-\boldsymbol{r},-t) \; ; \tag{2.3.71}$$

$$\psi(\boldsymbol{r},t) \to \gamma^5\psi(-\boldsymbol{r},-t), \quad \overline{\psi}(\boldsymbol{r},t) \to -\overline{\psi}(-\boldsymbol{r},-t)\gamma^5 \; . \tag{2.3.72}$$

The invariance of the field equations is easy to check. For example, the transformed Dirac equation reads:

$$i\frac{\partial}{\partial(-x_\mu)}\gamma_\mu\gamma_5\psi(-x) + \gamma^5 m\psi(-x) = 0 \; . \tag{2.3.73}$$

Multiplying by γ_5 on the left and using that $\{\gamma_\mu, \gamma_5\} = 0$ and $\gamma_5^2 = 1$, we obtain the original Dirac equation, which completes the proof. The invariance of the boson equations is trivial to check. However, when analysing the commutation relations for bosons, we find that they are not invariant, since the right hand side changes sign because $\Delta(x)$ is odd. When dealing with time reversal, we encountered a similar situation. We solved it, taking advantage of the occurrence of the imaginary unit in the commutation relations, and defined time reversal as an antilinear operator. Facing with this problem, Pauli postulated that the strong reflection shall produce a mapping that reverses the order of factors in the operator algebra. This does not affect the anticommutation relations of fermion fields.

Let us review our reasoning: we are seeking a transformation with the property (2.3.56), because then anything we can write for an interaction density would be automatically invariant under that set of operations. Now we will prove the CPT theorem in two steps: firstly, showing that a wide class of field theories are invariant under strong reflection; secondly, proving that the product of **C**, **P**, **T** taking in any order is identical to SR followed by Hermitian conjugation. We could convince ourselves that the change of order in SR is indeed essential, by looking at the Heisenberg equation:

$$i[P_0, f] = \frac{\partial f}{\partial x_0} \; , \quad P_0 = \int d^3x T_{00} \; , \tag{2.3.74}$$

where $T_{\mu\nu}$ is the energy-momentum tensor. Note that the time derivative changes sign, but the total energy P_0 should remain invariant. Hence, we must take $[P_0, f] \to [f, P_0]$ to preserve the original form.

The behaviour of boson fields

For densities constructed out of boson fields and derivatives (of finite order) of fields, we know that for each index we get a minus sign under SR. Change of order is irrelevant since the boson fields commute. In order to achieve the invariance we want, we need to postulate the usual connection between spin and statistics, so all bilinear forms of boson fields must be properly symmetrized according to Bose-Einstein statistics just as the bilinear covariants made of spinors must be antisymmetrized. Thus, for combinations of boson fields,

$$\mathcal{O}_{\mu\nu\ldots\lambda\sigma} \rightarrow (-1)^n \mathcal{O}(x)_{\mu\nu\ldots\lambda\sigma} \tag{2.3.75}$$

is satisfied. $\mathcal{L}$ and T_{00} are made up of these operators with indices contracted. With each contraction, the rank decreases by 2; hence:

$$\mathcal{L}(x) \rightarrow \mathcal{L}(-x) \ , \quad T_{00}(x) \rightarrow T_{00}(-x) \ , \tag{2.3.76}$$

as we would like.

The behaviour of fermion fields

The behavior of bilinear covariants of spinors is less obvious. To study them, we have to look into the transformations of scalars, vectors and tensors which can be formed by means of the γ-matrices. The transformation (2.3.72) followed by a change in the order of factors is the same as

$$\overline{\psi}_2 \Omega \psi_1 \rightarrow [(-\overline{\psi}_2 \gamma^5) \Omega (\gamma^5 \psi_1)]^T = -\psi_1^T \gamma^{5T} \Omega^T \gamma^{5T} \overline{\psi}_2^T \ , \tag{2.3.77}$$

because changing the order of factors causes $\overline{\psi}_\beta O_{\beta\alpha} \psi_\alpha \rightarrow \psi_\alpha O_{\beta\alpha} \overline{\psi}_\beta = \psi_\alpha (O^T)_{\alpha\beta} \overline{\psi}_\beta = (\overline{\psi} O \psi)^T$. Then, the usual connection between spin and statistics leads to the correct transformation properties:

$$: -\psi_1^T \left(\gamma^5 \Omega \gamma^5\right)^T \overline{\psi}_2^T : \ = \ : \overline{\psi}_2 \gamma^5 \Omega \gamma^5 \psi_1 : \tag{2.3.78}$$

We were thus led to introduce this additional postulate: *In the Lagrangian, all products are symmetrized with respect to Bose fields and antisymmetrized with respect to Fermi fields.* A postulate of this type was already required for the discussion of charge conjugation, so it naturally extends to the proof of the theorem. In the case of Dirac fields, the prescription is $\overline{\psi}_\alpha \Gamma_{\alpha\beta} \psi_\beta \rightarrow 1/2 \left[\overline{\psi}_\alpha \Gamma_{\alpha\beta} \psi_\beta - \psi_\beta \Gamma_{\alpha\beta} \overline{\psi}_\alpha\right]$. Hence, the two terms in equation (2.3.78) are equal. Here, Ω corresponds to the well-known combinations of γ matrices that appear in the spin one-half operators that we observe: the scalar (S) combination — Ω equal to the identity — which simply gives $\overline{\psi}\psi$; the vector (V) combination that appears in $i\overline{\psi}\gamma^\mu\psi$; the tensor ($T$) one, in the expression $\overline{\psi}\sigma_{\mu\nu}\psi$; the axial-vector ($A$) combination of γ matrices, identified with the term $i\overline{\psi}\gamma^\mu\gamma^5\psi$; and finally the pseudoscalar (P) term, identified with $i\overline{\psi}\gamma^5\psi$. It is trivial to check that

$$\gamma^5 \Omega \gamma^5 = (-1)^n \Omega \ , \tag{2.3.79}$$

where $n =$ even for S, T, P and $n =$ odd for V, A. Combining the results, we get

$$\overline{\psi}_2 \Omega \psi_1 \rightarrow (-1)^n \overline{\psi}_2 \Omega \psi_1 \ , \tag{2.3.80}$$

under SR. Without the normal spin-statistics relation, the result would be just the *opposite* of what we want.

We conclude that both $\mathcal{L}$ and T_{00} are invariant under strong reflection whether they are made up of boson fields, fermion fields or even combinations of the two, provided that the "normal" symmetrization relations hold.

2.3.2 Hermitian Conjugation

Strong reflection is defined on the operator algebra, corresponding to a mapping of the operator algebra into itself which reverses the order of factors in products, as legitimate mathematically as that which preserves the order. However, SR cannot be applied on a Hilbert space. Therefore, it was necessary to introduce a second mapping of the operator algebra into itself which also reverses products and which also leaves the interaction densities invariant. The result of the consecutive application of both operations is defined in a Hilbert space. This second operation can be identified with Hermitian conjugation (HC). It follows that a Hermitian Hamiltonian should also be required for the validity of this proof. If we then show that the consecutive application of both operations is equivalent to the product of **C**, **P** and **T**, we prove the invariance of the theory.

For boson fields, if we first apply SR and afterwards Hermitian conjugation, we end up with:

$$\phi(x) \xrightarrow{\text{SR}} \phi(-x) \xrightarrow{\text{HC}} \phi^\dagger(-x) \; ; \tag{2.3.81}$$

$$\phi_\mu(x) \xrightarrow{\text{SR}} -\phi_\mu(-x) \xrightarrow{\text{HC}} -\phi^\dagger_\mu(-x) \; . \tag{2.3.82}$$

Similarly, operating with **CPT**:

$$\phi(x) \xrightarrow{\text{C}} \alpha_C \phi^\dagger(x) \xrightarrow{\text{P}} \alpha_C \alpha_P^* \phi^\dagger(-\vec{x}, t) \xrightarrow{\text{T}} \alpha_C^* \alpha_P \alpha_T^* \phi^\dagger(-x) \; ; \tag{2.3.83}$$

$$\phi_\mu(x) \xrightarrow{\text{C}} \alpha'_C \phi^\dagger_\mu(x) \xrightarrow{\text{P}} \alpha'_C \alpha'^*_P \begin{cases} \phi^\dagger_0(-\vec{x}, t) \\ -\phi^\dagger_i(-\vec{x}, t) \end{cases} \xrightarrow{\text{T}} \alpha'^*_C \alpha'_P \alpha'^*_T \begin{cases} -\phi^\dagger_0(-x) \; . \\ -\phi^\dagger_i(-x) \; . \end{cases} \tag{2.3.84}$$

The two results are equal for $\alpha_C^* \alpha_P \alpha_T^* = 1$ and $\alpha'^*_C \alpha'_P \alpha'^*_T = 1$. For a different order of **C**, **P**, **T** operations, the condition on the phase may be modified.

Now we study the Dirac fields. We have:

$$\psi \xrightarrow{\text{SR}} \psi^T \gamma^{5T} \xrightarrow{\text{HC}} \gamma^5 \psi^{\dagger T} = \gamma^{5T} \gamma^{0T} \gamma^{0T} \psi^{\dagger T} = -\gamma^0 \gamma^5 \overline{\psi}^T \; . \tag{2.3.85}$$

While operating with CPT:

$$\begin{aligned} \psi \xrightarrow{\text{C}} \alpha''_C C \overline{\psi}^T \quad \xrightarrow{\text{P}} \quad & \alpha''_C \alpha''^*_P C \gamma^0 \overline{\psi}^T = -\alpha''_C \alpha''^*_P \gamma^0 C \overline{\psi}^T \xrightarrow{\text{T}} \\ & \alpha''^*_C \alpha''_P \alpha''^*_T \gamma^0 C T \overline{\psi}^T = \alpha''^*_C \alpha''_P \alpha''^*_T \gamma^0 \gamma^5 \overline{\psi}^T \; . \end{aligned} \tag{2.3.86}$$

Thus, CPT and SR followed by HC give the same result if $\alpha''^*_C \alpha''_P \alpha''^*_T = -1$. The condition on the phase is modified if **C**, **P** or **T** are applied in a different order. This completes the proof that a quantized field theory constructed from fields of spin $0, 1/2$ and 1 is invariant with respect to the product CPT taken in any order.

Our reasoning relied on the substitution of SR by an operator Θ (SR + HC) in Hilbert space (defined because the order of operators is reversed by SR but restored by HC), which does not change the physical observables. Then, under the combined operation, the theory remains invariant while its application is meaningful for both the operator algebra and for the underlying Hilbert space. We completed the proof demonstrating that Θ is indeed an antiunitary operator in Hilbert space: it is, up to a phase, identical to the product of three operators in Hilbert space **C**, **P** and **T**. The first two are unitary, while time reversal is antiunitary, hence Θ is an antiunitary operator. Note, finally, that the hermiticity of interaction densities was not necessary for proving the invariance under SR, but to identify CPT-operation with SR followed by HC.

2.3.3 Some comments on the proof

To prove the CPT Theorem, we followed the *proof* suggested by Lüders, that clarifies the physical ideas behind the successive application of these discrete symmetries, although it is not the most technical one, neither it includes the generalization for higher spin fields (Pauli gave a general proof using the theory of representations of the proper Lorentz group [6]).

We now summarize what we have assumed:

- Invariance under proper orthochronous Lorentz transformations;
- Interaction densities are local and constructed out of field operators and derivatives of field operators of finite order;
- The normal connection between spin and statistics (with kinematically independent Dirac fields anticommuting);
- Interaction densities are symmetrized and antisymmetrized in the proper way;
- The hermiticity of interaction densities.

Before continuing, we should realize the primary mathematical reasoning in what concerns discrete symmetry operations: we have a quantum field theory, which is invariant under proper Lorentz transformations, and we ask for additional symmetry properties which may hold in the theory. One has to state all the physical assumptions that the fields must satisfy, besides the transformation properties. So, before formulating the CPT Theorem, it is necessary to make a mathematical formulation of the operations **C**, **P** and **T** separately. At first sight, our task might consist on constructing all symmetry operations, which are either unitary or antiunitary — as it is stated in Wigner's Theorem — which transform the field operators $\phi_r(x)$ into $\phi_{r'}(x')$, with the spacetime coordinates being transformed by one of the four discrete Lorentz transformations: Identity, Space Reflection, Time Reversal or Inversion of all four coordinates. Having in mind the applications of these operations to interacting fields, we restrict our choices to local-symmetry operations, requiring that $\phi'_r(x')$ is connected with $\phi_r(x)$ by a linear transformation of the form:

$$\phi'_{r'}(x') = \sum_r \alpha_{r'r}\phi_r(x) \ . \tag{2.3.87}$$

With the additional supposition of locality, it is shown [6] that the symmetry group is generated by charge conjugation, space reflection and time-reversal.

2.4 Applications of the Theorem

Now, with the CPT Theorem in hand, we can deduce several interesting consequences for relativistic quantum field theory, in what concerns the relation between particles and antiparticles. The latter must exist, even if charge conjugation is not an exact symmetry of Nature. We focus on three predictions of the Theorem that we will use recurrently in the course of this work: (1) mass equality; (2) opposite sign of charge; and (3) equality of total decay widths of particle and antiparticle.

2.4.1 Mass equality

Consider a particle a at rest with the z-component of angular momentum m; then, applying the discrete operations C, P, T in some (irrelevant) order, we have:

$$TPC\,|a,m\rangle = TP\,|\bar{a},m\rangle = T\,|\bar{a},m\rangle = \alpha_{TCP}\,|\bar{a},-m\rangle \ , \tag{2.4.88}$$

where we have included a general (complex) phase, α_{CPT}, which corresponds to a combination of the phases that appear due to the successive application of each operation; $\bar{a}$ denotes the antiparticle. We next write:

$$\begin{aligned}(\text{mass})_a &= \langle a,m|H|a,m\rangle = \langle a,m|(TCP)^{-1}(TCP)|H|(TCP)^{-1}(TCP)|a,m\rangle \\ &= \langle \bar{a},-m|H|\bar{a},-m\rangle = (\text{mass})_{\bar{a}} \ ,\end{aligned} \tag{2.4.89}$$

where we have used the invariance of the Hamiltonian under CPT.

2.4.2 Opposite charges

If we apply each operation in CPT with no particular order, we conclude that the electromagnetic four-current transforms as $(CPT)j_\mu(x)(CPT)^{-1} = -j_\mu(-x)$. But $j_\mu(x) = (\rho(x), \mathbf{j})$, where ρ is the charge density; in particular, $(CPT)\rho(x)(CPT)^{-1} = -\rho(-x)$. Consequently, we have:

$$\begin{aligned}Q_a &= \langle a|\textstyle\int d^3x\rho(x)|a\rangle = \langle a|(CPT)^{-1}(CPT)\textstyle\int d^3x\rho(x)(CPT)^{-1}(CPT)|a\rangle \\ &= \langle\overline{a}|\textstyle\int d^3x[-\rho(-x)]|\overline{a}\rangle = -Q_{\overline{a}} \ .\end{aligned} \tag{2.4.90}$$

2.4.3 Equality of Lifetimes

If CPT symmetry is valid, the scattering matrix $\mathcal{S}$ transforms under the antilinear CPT transformation as [7]:

$$\mathcal{S} \to \mathcal{S}' = (CPT)^{\dagger}\mathcal{S}(CPT) = \mathcal{S}^{\dagger} \ . \tag{2.4.91}$$

In the quantum theory of interactions, the scattering matrix defines the amplitudes for finding the system — in the remote future — in the free[7] final state $|f\rangle$, when it was prepared to be in the free initial state $|i\rangle$ in the remote past, that is, $\mathcal{S}_{fi} \equiv \langle f_{(\mathrm{out})}|i_{(\mathrm{in})}\rangle$. These states form two complete sets of basis in the Hilbert space; or equivalently, $\mathcal{S}$ describes how a given 'in' state is expanded in terms of the 'out' states: $|i_{(\mathrm{in})}\rangle = \sum S_{fi}\,|f_{(\mathrm{out})}\rangle$, where the sum is extended to all possible final states. From this, the unitarity of the $\mathcal{S}$ matrix follows. In the interaction picture, the $\mathcal{S}$ operator is obtained as $\mathcal{S} \equiv U^{(I)}(\infty, -\infty)$ where U is the time evolution operator, defined through Heisenberg's equation: as the transformed operator must still satisfy the latter, in the new coordinates, it follows that $(CPT)^{\dagger}U^{(I)}(t,t_0)(CPT) = U^{I}(-t,-t_0)$; so that the scattering matrix is transformed into $S^{\dagger} = S^{-1}$, as 'in' and 'out' states are interchanged by the time reversal operation.

It is also usual to define $\mathcal{S}$ in terms of the transition matrix $\mathcal{T}$,

$$S_{fi} = \delta_{fi} - i(2\pi)\delta(E_f - E_i)\mathcal{T}_{fi} \ , \tag{2.4.92}$$

for which a similar relation holds:

$$\mathcal{T} \to \mathcal{T}' = (CPT)^{\dagger}\mathcal{T}(CPT) = \mathcal{T}^{\dagger} \ . \tag{2.4.93}$$

Given these brief considerations, one can consider the lifetime of a particle $|a\rangle$ which can decay into states $|f\rangle$ due to the effect of interactions H_{int}:

$$\Gamma_a \equiv \tau_a^{-1} \propto \sum_f |\langle f|\mathcal{T}|a\rangle|^2 \ . \tag{2.4.94}$$

Although a phase space factor is included in the calculation of the total decay width, as given by *Fermi's Golden Rule*, it is the same for particles and antiparticles, because of the equality of their masses. Hence, the only concern is the transition matrix; for this

$$\sum_f |\langle f|\mathcal{T}|a\rangle|^2 = \sum_f \langle a|\mathcal{T}^{\dagger}|f\rangle\langle f|\mathcal{T}|a\rangle = \langle a|\mathcal{T}^{\dagger}\mathcal{T}|a\rangle \ , \tag{2.4.95}$$

due to the completeness of the sum over the final states. Using the results from CPT invariance (2.4.91), we get:

$$\langle a|\mathcal{T}^{\dagger}\mathcal{T}|a\rangle = \langle a|(CPT)^{\dagger}\mathcal{T}\mathcal{T}^{\dagger}(CPT)|a\rangle = \langle \bar{a}_{PT}|\mathcal{T}\mathcal{T}^{\dagger}|\bar{a}_{PT}\rangle = \langle \bar{a}_{PT}|\mathcal{T}^{\dagger}\mathcal{T}|\bar{a}_{PT}\rangle \ , \tag{2.4.96}$$

[7]The so-called 'in' and 'out' states are eigenstates of the full Hamiltonian ($H_0 + H_{\mathrm{int}}$), which specify the particle content at time $t = -\infty$ and $t = +\infty$, respectively ($H_{\mathrm{int}} \to 0$ for $t \pm \infty$).

where the last operation $\mathcal{T}\mathcal{T}^\dagger = \mathcal{T}^\dagger\mathcal{T}$ results from the unitary of $\mathcal{S}$[8].

We now insert a sum over a complete set of states, yielding:

$$\sum_f \langle \bar{a}_{PT}|\mathcal{T}^\dagger|f\rangle \langle f|\mathcal{T}|\bar{a}_{PT}\rangle = \sum_f \langle f|\mathcal{T}|\bar{a}_{PT}\rangle^* \langle f|\mathcal{T}|\bar{a}_{PT}\rangle = \sum_f |\langle f|\mathcal{T}|\bar{a}_{PT}\rangle|^2 \; . \tag{2.4.97}$$

Since the decay width cannot depend on the spin orientation (which is the effect of PT) due to rotational invariance, the result follows:

$$\Gamma_a = \Gamma_{\bar{a}} \; . \tag{2.4.98}$$

From the proof it does not follow that the partial widths of various decay channels must be identical for particles and antiparticles; this is not required by the CPT Theorem (as it is by the C or CP symmetries[9]). The laws $m_a = m_{\bar{a}}$ and $\Gamma_a = \Gamma_{\bar{a}}$, and thus the validity of the CPT Theorem, have been checked experimentally with high precision. The most sensitive probe for this purpose is the neutral K meson — the physics of this system is discussed in the second chapter of Baym's book [8]. Since the decays of K^0 and its antiparticle $\bar{K}^0$ interfere (the physical particles are combinations of these two states), an upper bound for the mass difference is obtained with exceedingly high precision [9]:

$$m_{K^0} - m_{\bar{K}^0} < 4.0 \times 10^{-19}\ \text{GeV} \; . \tag{2.4.99}$$

A typical result for the agreement of particle and antiparticle lifetimes has been obtained for muons [9]:

$$\left|\frac{\tau_{\mu^+} - \tau_{\mu^-}}{\tau_{average}}\right| < (2 \pm 8) \times 10^{-5} \; . \tag{2.4.100}$$

2.5 (Good) Limitations of the CPT Invariance

As we have been discussing, while CPT invariance guarantees the equalities of particle and antiparticle masses and total decay widths, it does not require the partial decay widths for particle and antiparticle to be equal. This is of fundamental importance in what concerns the present baryon asymmetry of the Universe.

Let us contextualize the various work concerning CPT symmetry. Between 1955 and 1957, Lüders and Pauli derived explicit proofs of the CPT Theorem [6], based essentially on the operation of strong reflection (in the literature, the theorem is sometimes called the Lüders-Pauli Theorem). In another paper published in 1957, Lüders and Zumino investigated what are the connections between properties of particles and antiparticles that follow from the general CPT invariance [10]. The fact that it permits unequal partial decay widths was emphasized by Okubo [11], in the same year. He investigated the two

[8]Using the unitarity of $S = 1 - iT$, it follows that $1 = SS^\dagger = 1 - iT + iT^\dagger + TT^\dagger$. Similarly, $1 = S^\dagger S = 1 + iT^\dagger - iT + T^\dagger T$; thus, we conclude that $TT^\dagger = T^\dagger T$.

[9]The processes $i \to f$ and $\bar{i} \to \bar{f}$ have the same transition amplitude if charge conjugation is a symmetry of the theory; the same is true for $i \to f$ and $f_T \to i_T$ if this applies to time reversal (or, in the case of spinless particles, $\Gamma[i \to f] = \Gamma[f \to i]$)

decay modes of the Σ^+ hyperon and the corresponding ones for its antiparticle:

$$\Sigma^+ \to p + \pi^0 \ ; \tag{2.5.101}$$

$$\Sigma^+ \to n + \pi^+ \ ; \tag{2.5.102}$$

$$\overline{\Sigma}^+ \to \overline{p} + \pi^0 \ ; \tag{2.5.103}$$

$$\overline{\Sigma}^+ \to \overline{n} + \pi^- \ . \tag{2.5.104}$$

As a result of the CPT Theorem, the total lifetimes of $\overline{\Sigma}^+$ and Σ^+ are equal. However, computing the ratio between the relative frequency of the events (2.5.101) and (2.5.103), he found out it can strongly deviate from unity, if charge conjugation or time reversal does not hold (as in the case of weak interactions), meaning the branching ratios can be different for particle and antiparticle. Going back to section 2.4.3, it should be obvious why the proof does not extend to partial decay widths: unlike CP, CPT inverts the in and out states; so, the sum over a complete set of final states is crucial to obtain the equality (2.4.98).

Suppose that the Σ^+ flux is bigger than the $\overline{\Sigma}^+$ flux: this can produce p faster than $\overline{p}$. CP violation in the decay producing the hyperons can actually produce a charge asymmetry to get the Σ^+ flux bigger than the other. This leaves a hint that CPT can be satisfied, CP violated and a baryon asymmetry can arise, if $\overline{\Sigma}^+$ and Σ^+ are produced equally, $\overline{p}$ and p are produced equally, but $\overline{\Sigma}^+$ decays *less* in a mode that produces $\overline{p}$ and Σ^+ decays *more* in a mode that produces p.

Later on, Sakharov (1967) proposed the three necessary conditions for baryogenesis [12], building on Okubo's work.

2.6 CPT Violation in the Early Universe

The CPT Theorem has withstood numerous high-precision experimental tests, one of the sharpest quoted by the Particle Data Group involving the kaon particle-antiparticle mass difference (2.4.99). Given this experimental precision and since CPT Theorem holds generally for relativistic particle theories, any sign of its violation would be the signature of unconventional physics. It is thus a field of interest to examine possible theoretical mechanisms through which this invariance could be broken.

Obviously, an immediate possibility to test CPT-violating effects is to construct a theory disobeying any of the assumptions that enter the theorem. Another approach is to go beyond the Standard Model of Particle Physics, considering — for example — string theories. Without entering into details, we can readily be convinced that the usual premises of the theorem (e.g. locality) might be altered considering that strings are extended objects.

2.6.1 Spontaneous CPT Violation

Being Lorentz invariance one of the major axioms leading to the CPT Theorem, the relation between the breakdown of these two symmetries has been widely discussed in the literature. Moreover, it has been shown [13] that — in some string theories — CPT violation may, in fact, occur through a mechanism based on the spontaneous breaking of Lorentz symmetry, which may lead to observable effects at the current energy levels accessible for experiments. A natural way for this to occur is imagining that a higher-dimensional action, which is Lorentz and CPT invariant, exists in Nature. Then, the higher-dimensional Lorentz group would have to be spontaneously broken, in order to describe our four-dimensional world: this could in principle induce spontaneous CPT breaking.

Spontaneous Lorentz violation can occur in string theory because interactions can trigger nonzero expectation values for Lorentz-tensor fields — which do not appear in our renormalizable gauge theories in four dimensions. But the mechanism of spontaneous symmetry violation is well established in various fields, like Condensed Matter Physics and Particle Physics; it is also very attractive because the symmetry is violated through non-trivial ground-state solutions, while the underlying dynamics of the system remains invariant. For example, consider classical electrodynamics: the energy density associated to some electromagnetic field configuration is $V(\overrightarrow{E}, \overrightarrow{B}) = \frac{1}{2}(\overrightarrow{E}^2 + \overrightarrow{B}^2)$. The ground state is usually identified with the lowest-energy configuration of a system; in this case, this corresponds to the field-free one, so the vacuum is empty.

Let us now think about the Higgs field, which is a scalar. According to the mechanism that explains how particles gain mass, the Higgs potential is identified with $V(\phi) = (\phi^2 - \lambda^2)^2$ where λ is a constant. Again, the lowest possible field energy is zero. Now this requires, however, ϕ to be non-vanishing: $\phi^2 = \lambda^2$. It follows that the vacuum of a system containing a Higgs-type field is not empty, but instead contains the constant scalar field $\phi_{vac} \equiv \langle\phi\rangle = \lambda e^{i\theta}$, which is called the vacuum expectation value (VEV) of the field. Note that, after expanding around the VEV, quadratic terms on SM fields appear in the Lagrangian, so they develop mass. This is the Higgs mechanism at work. Also, note that $\langle\phi\rangle$ — being a scalar — does not choose a preferred direction in spacetime.

Finally, we take a look at a vector field $\overrightarrow{C}$, whose existence is not predicted within the framework of the SM. In analogy with the Higgs case, we take the expression for the energy density to be $V(\overrightarrow{C}) = (\overrightarrow{C}^2 - \lambda^2)^2$. Just like in the previous examples, this requires a non-vanishing VEV: $\overrightarrow{C}_{vac} \equiv \langle\overrightarrow{C}\rangle = \overrightarrow{\lambda}$. In this case, however, the true vacuum state contains an intrinsic direction, violating rotation invariance and thus Lorentz symmetry. Interactions leading to this type of phenomena can be found in the context of strings. Alan Kostelecký was the main driver behind the effort to develop a conceptual framework and procedure for treating spontaneous CPT and Lorentz violation [14]. It is assumed that underlying the effective four-dimensional action is a complete fundamental theory based on conventional quantum physics and that is CPT and Poincaré invariant. The fundamental theory is assumed to undergo spontaneous CPT and Lorentz breaking. We remark some of the ideas concerning this theoretical approach.

2.6.2 CPT Violation in String Theory

Spontaneous breaking of Lorentz symmetry can occur in a theory that contains certain types of interaction among Lorentz tensor fields, if such interactions produce nontrivial VEVs.

In string theory, solutions exist in which scalar field components have a nonzero value. This can lead to an effective action for tensor field components that would give a VEV to the latter. For example, the bosonic section of string field theory [15] contains the three field interaction term $A_\mu A^\mu \phi$, between the scalar ϕ and the massless vector A_μ. It follows that if the scalar field acquires a VEV, this would contribute in turn to a squared mass for A_μ. Hence, for the appropriate sign of $\langle\phi\rangle$, A_μ could also get a vacuum expectation value, breaking Lorentz invariance. As A_μ has one uncontracted index, the low-effective interaction term $A_\mu j^\mu$ is odd under CPT. In this framework, A_μ is viewed as a background field permeating the spacetime vacuum. In a CPT transformed version of an experiment, this will not change, but the generalized current will, leading to effective CPT violation, as well as Lorentz violation. The behaviour of these expectation values as background fields will be explained further on. We remark the importance of identifying CPT with the Strong Reflection operation, to recognize the CPT-violating terms that could appear in the action.

2.6.3 Standard Model Extension

Starting from the conventional SM Lagrangian, Lorentz-breaking modifications $\mathcal{L}'$ can be added in a simple way:

$$\mathcal{L}_{\mathrm{SME}} = \mathcal{L}_{\mathrm{SM}} + \mathcal{L}' \ , \tag{2.6.105}$$

where the subscript SME refers to the generalized theory designating the Standard Model Extension. Then, spontaneous CPT violation arises from nonzero expectation values acquired by some Lorentz tensor T. Any interaction that is part of a four-dimensional effective theory must have mass dimension four. M is considered to be the scale of the high momenta, that is, the scale of unification, which is large compared to the scale m of the effective theory. The expectation values $\langle T\rangle$ of the tensors T are assumed to be Lorentz and possibly CPT violating, so any terms that survive after the spontaneous symmetry breaking and are contemplated in $\mathcal{L}'$ must be suppressed by, at least, on power of m/M relative to the scale of the effective theory. As an example, consider the schematic form of terms that can appear in the fermionic sector of the low-energy limit of the underlying theory:

$$\mathcal{L}' \supset \frac{\lambda}{M^k}\langle T\rangle \cdot \overline{\psi}(\Gamma)(i\partial)^k\chi + h.c. \ , \tag{2.6.106}$$

where ψ, χ are fermion fields, λ is a dimensionless coupling constant and Γ some gamma-matrix structure. The procedure extends to add to the Lagrangian all possible extra terms that can incorporate the effects of spontaneous Lorentz and CPT breaking at the level of the SM. This set is restricted by allowing only hermitian terms that preserve $SU(3)\times SU(2)\times U(1)$ gauge invariance and power-counting renormalizability in the extended action. Following these requirements, a general Lorentz-violating extension of the

SM that includes both CPT-even and CPT-odd terms has been constructed [16].

Since, in this context, CPT violation arises from nonzero expectations of Lorentz tensors, Lorentz invariance is necessarily spontaneously violated too. The converse is false: expectation values of Lorentz tensors with an even number of indices preserve CPT. For example, for the case $k = 0$, there are two possible types of CPT-violating bilinears:

$$\mathcal{L}'_a = a_\mu \overline{\psi} \gamma^\mu \psi, \quad \mathcal{L}'_b = b_\mu \overline{\psi} \gamma_5 \gamma^\mu \psi \tag{2.6.107}$$

where a_μ, b_μ are interpreted as effective couplings arising from the spontaneous symmetry breaking which are invariant under CPT transformations. Being interpreted as background fields, these tensorial coefficients cause the terms in (2.6.107) to break CPT (having one Lorentz-index). On the contrary, it is clear why the bilinears involving $\sigma_{\mu\nu}$ and γ_5 separately do not break CPT invariance. For completeness, we also write the most relevant terms appearing in the case $k = 1$:

$$\mathcal{L}'_c = \frac{1}{2} i c^\alpha \overline{\psi} \overleftrightarrow{\partial_\alpha} \psi \ , \quad \mathcal{L}'_d = \frac{1}{2} d^\alpha \overline{\psi} \gamma^5 \overleftrightarrow{\partial_\alpha} \psi \ , \quad \mathcal{L}'_e = i e^\alpha_{\mu\nu} \overline{\psi} \sigma^{\mu\nu} \overleftrightarrow{\partial_\alpha} \psi \ , \tag{2.6.108}$$

which, in turn, break CPT as well. Here $A \overleftrightarrow{\partial_\mu} B \equiv A \partial_\mu B - (\partial_\mu A) B$. In all these expressions, the quantities $a_\mu, b_\mu, c^\alpha, d^\alpha$ and $e^\alpha_{\mu\nu}$ (which are real due to the presumed hermiticity of the underlying theory) are combinations of tensor VEVs, coupling constants, mass parameters and coefficients arising from the decomposition of Γ.

In this framework, the violation of CPT invariance gives rise to the possibility of Baryogenesis in the early Universe, as reported in Ref. [17]. This is possible, identifying a chemical potential in the interaction (2.6.106) and constraining the free parameters (the cutoff scale and the decoupling temperature) to give rise to the observed baryon asymmetry.

2.6.4 Background Fields

Next, we want to understand how Lorentz invariance is broken by these type of terms. Not only the effective theory, but also the underlying one are constructed to be explicitly Lorentz invariant. All the interaction densities in (2.6.107) and (2.6.108) have contracted indexes, that is, they are coordinate Lorentz scalars. By construction, the SME extension is thus invariant under rotations or boosts of an observer's inertial frame - these are called *observer* Lorentz transformations. They should be contrasted with rotations or boosts of the localized fields in a fixed observer coordinate system, called *particle* Lorentz transformations.

The distinction between observer and particle transformations is essential to understand the breakdown of CPT symmetry in the present model. The CPT-violating terms are interpreted as arising from constant background fields, like a_μ and b_μ. These eight quantities (four components for each field) transform as two 4-vectors under observer Lorentz transformations and as eight scalars under particle Lorentz transformation, whereas they are coupled to currents that transform as 4-vectors under both types of transformation. Hence, observer Lorentz symmetry is still an invariance of the model, but the *particle Lorentz group* is broken.

As an example, to understand the role of the so-called background fields, consider the following analogy [16]: an electron with momentum perpendicular to a uniform magnetic field moves in a circle. Suppose that, in the same observer frame, we instantaneously increase the magnitude of the electron velocity without changing its direction, causing the electron to move in a circle of larger radius. This particle boost is an observable effect in that particular frame, but leaves the background field unaffected. However, if instead we apply an observer boost perpendicular to the magnetic field, the electron no longer moves in a circle. In the new inertial frame, this is viewed as an $\overrightarrow{E} \times \overrightarrow{B}$ drift. The background magnetic field is thus transformed into a new electromagnetic field under observer boosts, but it is unchanged by particle boosts. This scalar character of the background field under this type of Lorentz transformations that change the physics in a fixed inertial frame is what defines a particle transformation.

In this model, the behavior of the expectation fields as backgrounds is a consequence of their origin as nonzero VEVs of Lorentz tensors in the bare theory. They are a global feature of the effective model and cannot arise from localized experimental conditions. The key aspect for generating Lorentz- (and CPT-) violating terms while preserving observer Lorentz invariance (or coordinate independence) is the spontaneous Lorentz breaking in an underlying fully Lorentz-invariant theory, like string theory. Since observer Lorentz invariance constrains the physics under coordinate changes made by an external observer, an underlying theory with this property cannot lose it through internal interactions such as those leading to spontaneous Lorentz violation.

However, this presents an immediate consequence: measurements of the expectation coefficients will give different results in different observer's frames. For example, considering a particle-antiparticle oscillating system, like the neutral meson system, a Lagrangian term as (2.6.107) would introduce a correction of the form $\Delta\Lambda \approx \overrightarrow{p} \cdot \overrightarrow{a}$, where $\overrightarrow{p}$ is the direction of the particle beam and $\Delta\Lambda$ is the difference of the diagonal terms in the Hamiltonian of the system (which, if non-zero, directly breaks CPT invariance). As the Earth rotates, the direction of the experiment changes because it is attached to the Earth, thus leading to a possible observable effect [18].

Chapter 3

Thermal History

In this chapter, we briefly discuss the first minutes in the history of the Universe.

Since we will use observational parameters to restrict the variables of our model, we need to understand what are the cosmological evidences for a non-vanishing asymmetry and which parameters that we measure today can be used to fix it in the first stages of evolution.

Therefore, it is crucial to review also the basic dynamics of an expanding Universe at early times, which will lead us to conclude that, providing the Universe is isentropic (as in General Relativity), this asymmetry is deduced from the observed baryon-to-entropy ratio, $\eta_S \sim 10^{-10}$. This parameter is determined by precise WMAP and Planck measurements of the CMB anisotropy spectrum and the predictions for the light element abundances produced during Big Bang Nucleosynthesis, meaning that Baryogenesis is strictly constrained by these values. We explore some important features relative to BBN, namely the freeze-out of weak interactions, which sets the starting point for the formation of primordial elements, and how these are sensitive to the high entropy of the Universe ($\eta_S \ll 1$).

The high precision concordance of these two measurements (CMB+BBN) is one of the greatest successes in Cosmology. Hence, we also explore this link, confronting observational data and showing that the observed asymmetry cannot be explained in the framework of the SM.

All this leads to the conclusion that some mechanism must be in place in order to generate the overabundance of baryons compared to antibaryons.

3.1 Standard Cosmology

The observed Universe is composed of three components, each dominating at different epochs in the expansion history due to their different dilution rates. They are: cosmological constant Λ with a constant energy density (although more complicated proposals abound [19]); radiation (diluted as a^{-4}); and non-relativistic matter with $p \ll \rho$ (diluted slower, as a^{-3}). This means that radiation dominated the very early Universe; later on, heavy dark and baryonic matter became relevant; and finally — up to the present moment — Λ has been governing the dynamics.

3.1.1 Basic Concepts in General Relativity

Cosmological dynamics follow from the Einstein equation,

$$G_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = 8\pi G T_{\mu\nu} \ , \tag{3.1.1}$$

where $G_{\mu\nu}$ is the Einstein tensor, $R_{\mu\nu}$ is the Ricci tensor, R is the Ricci scalar, $g_{\mu\nu}$ is the metric, G is Newton's gravitational constant and $T_{\mu\nu}$ is the stress-energy tensor (here we are neglecting the cosmological constant).

Admitting the homogeneity and isotropy of space, Einstein equation is simplified with the Robertson-Walker metric, which yields the line element written in comoving coordinates as

$$ds^2 = dt^2 - a^2(t) \frac{dr^2}{1 - kr^2} + r^2 d\Omega^2 \ , \tag{3.1.2}$$

where $a(t)$ is the cosmic scale factor and k the curvature.

Matter is assumed to behave as a perfect fluid, with an energy-momentum tensor

$$T^{\mu\nu} = (p + \rho) u^\mu u^\nu - p g^{\mu\nu} \ , \tag{3.1.3}$$

where ρ and p are the energy density and pressure of the fluid; and u^μ is its four-velocity, normalized so that $u^\mu u_\mu = 1$.

In GR, the geometry of spacetime is determined by the metric tensor and the metric connection, which defines the covariant derivative on the spacetime. The former is written in terms of the Christoffel symbols:

$$\Gamma^\sigma_{\mu\nu} = \frac{1}{2} g^{\sigma\alpha} (g_{\mu\alpha,\nu} + g_{\alpha\nu,\mu} - g_{\mu\nu,\alpha}) \ , \tag{3.1.4}$$

which are the general form of a metric-compatible and torsion-free affine connection, in General Relativity. This allows us to define the Riemann curvature tensor:

$$R^\rho_{\sigma\mu\nu} = \Gamma^\rho_{\nu\sigma,\mu} - \Gamma^\rho_{\mu\sigma,\nu} + \Gamma^\rho_{\mu\lambda}\Gamma^\lambda_{\nu\sigma} - \Gamma^\rho_{\nu\lambda}\Gamma^\lambda_{\mu\sigma} \ . \tag{3.1.5}$$

From it, one defines the Ricci tensor, $R_{\mu\nu} = R^\rho_{\mu\rho\nu}$, which, upon contraction with the metric, gives the Ricci scalar, $R = g^{\mu\nu} R_{\mu\nu}$.

Ploughing through the amount of algebra required to calculate the non-zero components of the connection, it is then straightforward to compute the Ricci scalar [20]:

$$R = -6 \left(\frac{\ddot{a}}{a} + \frac{\dot{a}^2}{a^2} + \frac{k}{a^2} \right) \ . \tag{3.1.6}$$

The $\mu = \nu = 0$ component of Einstein's equation gives the Friedmann equation:

$$H^2 = \frac{8\pi G}{3} \rho \ , \tag{3.1.7}$$

defining the Hubble parameter, $H \equiv \dot{a}/a$. We set $k = 0$, according to strong evidence pointing towards a flat Universe; this comes from the fact that the sum of the observed abundances — $\Omega = \rho/\rho_{crit} \equiv 3H_0^2\rho/8\pi G$, where H_0 is the present value of the Hubble parameter — for each cosmological component is remarkably close to unity.

Likewise, the $\mu = i, \nu = j$ components yield

$$-2\left(\frac{\ddot{a}}{a} + \frac{\dot{a}^2}{a^2}\right) = 8\pi G p \ . \tag{3.1.8}$$

The quantity $G_{\mu\nu}$ is covariantly conserved in GR, $G^{\mu\nu}_{;\nu}$=0 (the *proof* relies on the Bianchi identities for the Riemann tensor). This implies, using Einstein equation, that the energy momentum tensor (3.1.3) is also covariantly conserved, reflecting energy and momentum conservation in an expanding Universe:

$$\begin{aligned} T^{0\nu}_{;\nu} &= T^{0\nu}_{,\nu} + \Gamma^0_{\alpha\nu}T^{\alpha\nu} + \Gamma^\nu_{\alpha\nu}T^{0\alpha} \\ &= T^{00}_{,0} + \Gamma^0_{ij}T^{ij} + \Gamma^\nu_{0\nu}T^{00} \\ &= \dot{\rho} + 3\frac{\dot{a}}{a}(\rho + p) = 0 \ . \end{aligned} \tag{3.1.9}$$

The Friedmann equation (3.1.7) admits not only non-relativistic dust, but also other types of matter and energy. In order to investigate how different fluids evolve in an expanding Universe, we can start by considering the First Law of Thermodynamics,

$$dE = -pdV + TdS \ , \tag{3.1.10}$$

where the variation of energy in a given volume is due to the work or heat transferred in and out of it.

If the expansion of the Universe was adiabatic, we would set $TdS = 0$. Then, for a fluid of energy density ρ and pressure p enclosed in a comoving volume $V = a^3$, we would have:

$$d(\rho a^3) = -pd(a^3) \ . \tag{3.1.11}$$

Differentiating with respect to time, one can write the equation above as a conservation law:

$$\dot{\rho} = -3H(\rho + p) \ . \tag{3.1.12}$$

As this gives the same result as equation (3.1.9), our Universe can —indeed — be regarded as isentropic[1]. Equation (3.1.12) also let us know the evolution of the energy density: for a simple equation of state, $p = \omega\rho$, it evolves like $\rho \propto a^{-3(1+\omega)}$. Examples of interest include dust/ nonrelativistic matter ($\omega = 0$, so that $\rho_m \propto a^{-3}$), radiation ($\omega = 1/3$, implying $\rho_r \propto a^{-4}$) and a cosmological constant ($\omega = -1$, which means $p_\lambda = -\rho_\lambda$).

3.1.2 Expansion Rate, Decoupling of Species and Freeze-out

The key to understand the thermal history of the Universe is the comparison between the rate of interactions Γ of the relevant processes and the rate of the expansion H. This can be done by looking at

[1]Note that we have assumed isentropy in the form of $T^{\mu\nu}$ as a perfect fluid.

the relevant terms in the equation describing the evolution of the particle number density, with a collision term taking into account the interactions with different particles. This corresponds to the general Boltzmann equation:

$$\dot{n}_i + 3Hn_i = -\langle\sigma_i v_i\rangle \left[n_i^2 - (n_i^{\text{eq}})^2\right] \ , \tag{3.1.13}$$

where n_i^{eq} denotes the equilibrium distribution of the particle species, σ_i is the total cross section of processes that either destroy or create them, and v_i is their interaction velocity. We can identify the proportionality constant with the particle interaction rate,

$$\Gamma \equiv n\sigma v \ . \tag{3.1.14}$$

Then, equation (3.1.13) can be recast in the form

$$\frac{d\ln N_i}{d\ln a} = -\frac{\Gamma}{H}\left[1 - \left(\frac{N_i^{\text{eq}}}{N_i}\right)^2\right] \ , \tag{3.1.15}$$

showing that any deviation from thermal equilibrium makes the system evolve towards thermal equilibrium once again. For example, starting with a deficit of particles, $N_i \ll N_i^{\text{eq}}$, the r.h.s. gives a positive contribution, making the number of particles grow until its equilibrium value. Similarly, if the number of particles exceeds its equilibrium value, more particles will be destroyed until equilibrium is achieved.

When $\Gamma \gg H$, the collision term dominates and the system quickly relaxes to a steady state with particles assuming their equilibrium abundances. In other words, the timescale of particle interaction is much smaller than the characteristic expansion time scale,

$$t_C \equiv \frac{1}{\Gamma} \ll t_H \equiv \frac{1}{H} \ , \tag{3.1.16}$$

meaning that local equilibrium is reached before expansion becomes relevant. As the Universe cools, however, interactions might not be able to keep up with the Hubble expansion (t_H becomes larger): at the critical time, $t_C \sim t_H$, particles decouple from the thermal bath.

When $\Gamma < H$, the collision term cannot compensate for the Hubble expansion and the system departs from thermal equilibrium. This implies that the abundance of the particle species i will remain constant after the decoupling, which is known as freeze-out.

In the primitive Universe, for temperatures $T \gg 10^2$ GeV, all the Standard Model (SM) particles behaved as radiation, dominating the energy density of the Universe. We want to establish the magnitude of the ratio Γ/H for scattering processes mediated by gauge bosons, which depends on whether they are relativistic or non-relativistic at the decoupling time. Defining the generalized structure constant $\alpha \equiv g_A^2/4\pi$ for the gauge coupling associated with a generic gauge boson A, we have[2]:

[2] In both limits, the cross section must be proportional to α^2, since we are considering processes corresponding to diagrams with two vertices involved. In the case of a massless boson, we cannot use the mass to describe the order of magnitude of a typical cross section; instead, we use $\sigma \sim L^2 \sim T^{-2}$. In the case of a massive boson, the latter becomes $(T/m)^2$, as obtained rigorously in Sec. 3.5.

- For a massless boson, $m_A \ll T$, the cross section is $\sigma \propto \frac{\alpha^2}{T^2}$;
- For a massive gauge boson[3], $m_A \gtrsim T$, the typical cross section is $\sigma \sim G_A^2 T^2$, where $G_A \sim \alpha/m_A^2$ is the generalized Fermi constant.

From dimensional analysis, $\Gamma \sim \sigma T^3$, as for relativistic particles $v \sim 1$; while, from the Friedmann equation (3.1.7), we get $H \sim T^2/m_P$, where $m_P = 1/\sqrt{G}$ is the Planck mass. Hence, the two conditions above translate to (up to constant factors)

$$\frac{\Gamma}{H} \sim \begin{cases} \alpha^2 m_P/T \ , & \text{for} \quad m_A \ll T \\ G_A^2 m_P T^3 \ , & \text{for} \quad m_A \gtrsim T \ . \end{cases} \tag{3.1.17}$$

This means that a particle species can be in thermal equilibrium at high temperature, $m_A \ll T \ll m_P$; and decouple later on from the thermal bath, once the temperature drops below $T_D \sim (G_A^2 m_P)^{-1/3} \sim \alpha^{-2/3}(m_A/m_P)^{1/3} m_A \lesssim m_A$. For species which interact only through weak interactions ($m_W \sim m_Z \sim 80$ GeV, corresponding to $G_F \sim 10^{-5}$ GeV^{-2}), the decoupling temperature is near $T \sim 1$ MeV.

If expansion had never overcome the interaction rates, the Universe would be filled with mostly photons and equilibrium would make any other type of matter be destroyed at the same rate it is created. Hence, it is crucial to understand the deviations from equilibrium that led to the freeze-out of massive particles.

3.2 Equilibrium Thermodynamics

As we have seen, the SM predicts that, in the very early Universe, most particles were in thermal equilibrium with photons — a *thermal bath* of particles. In order to describe this state and the subsequent evolution, we need to review the basic properties of particle distributions in kinetic thermal equilibrium [20]. For that, we use the familiar Bose-Einstein or Fermi-Dirac distributions which give to a system of particles a level occupancy f, at a temperature T:

$$f(\mathbf{p}) = \frac{1}{e^{\frac{E-\mu}{T}} \pm 1} \ , \tag{3.2.18}$$

where $E^2 = |\mathbf{p}|^2 + m^2$ is the energy of particles; $\mu = (\rho + p)/n$ is the chemical potential; the $(+)$ sign corresponds to fermions, while the $(-)$ refers to bosons. Local thermal equilibrium means not only kinetic equilibrium, but also chemical equilibrium. Then if, for example, a species A interacts with species B, C and D via scattering processes such as

$$A + B \longleftrightarrow C + D \ , \tag{3.2.19}$$

chemical equilibrium implies that $\mu_A + \mu_B = \mu_C + \mu_D$.

[3]Examples of these are $W^\pm$ or Z, the gauge bosons of the weak interaction, which receive masses below the scale of electroweak symmetry breaking ($T \lesssim 10^2$ GeV).

Recalling statistical mechanics, we know that the density of states for a particle with g internal degrees of freedom (e.g., spin) is $g/(2\pi)^3$; in this way, we can compute the particle density in phase space as the density of states times the distribution function. The phase space distribution allows us to evaluate the associated number density n, energy density ρ and pressure p for our particle system:

$$n = g \int \frac{d^3p}{(2\pi)^3} f(\mathbf{p}) \ , \tag{3.2.20}$$

$$\rho = g \int \frac{d^3p}{(2\pi)^3} E(\mathbf{p}) f(\mathbf{p}) \ , \tag{3.2.21}$$

$$p = g \int \frac{d^3p}{(2\pi)^3} \frac{|\mathbf{p}|^2}{3E(\mathbf{p})} f(\mathbf{p}) \ . \tag{3.2.22}$$

Next, we analyse the asymptotic limits of these expressions (assuming, for now, $|\mu| \ll T$).

Relativistic species

For $T \gg m$, the Bose-Einstein or Fermi-Dirac distribution becomes

$$f(y) = \frac{1}{e^y \pm 1} \tag{3.2.23}$$

where $y = |\mathbf{p}|/T$. This implies, for the particle number density,

$$n = g \int_0^{+\infty} \frac{4\pi |\mathbf{p}|^2 d|\mathbf{p}|}{(2\pi)^3} \frac{1}{e^y \pm 1} = \frac{g}{2\pi^2} T^3 \int_0^{+\infty} \frac{y^2}{e^y \pm 1} dy \ . \tag{3.2.24}$$

To evaluate integrals like this one, it is useful to consider the following results

$$\frac{1}{e^x + 1} = \frac{1}{e^x - 1} - \frac{2}{e^{2x} - 1} \ ; \tag{3.2.25}$$

$$\int_0^{+\infty} \frac{y^n}{e^y - 1} dy = \zeta(n+1)\Gamma(n+1) \ , \tag{3.2.26}$$

where $\zeta(z)$ is the Riemann zeta-function.

For bosons, it is easy to obtain

$$n_b = \frac{g}{\pi^2} T^3 \zeta(3) \ . \tag{3.2.27}$$

For fermions, we need to develop the integrand:

$$\begin{aligned} n_f &= \frac{g}{2\pi^2} T^3 \left(\int_0^{+\infty} \frac{y^2}{e^y - 1} dy - 2 \int_0^{+\infty} \frac{y^2}{e^{2y} - 1} dy \right) \\ &= \frac{g}{2\pi^2} T^3 \left(\int_0^{+\infty} \frac{y^2}{e^y - 1} dy - \frac{1}{4} \int_0^{+\infty} \frac{z^2}{e^z - 1} dz \right) \\ &= \frac{g}{2\pi^2} T^3 \left(\zeta(3)(2!) - \frac{1}{4} \zeta(3)(2!) \right) \\ &= \frac{3}{4} \frac{g}{\pi^2} \zeta(3) T^3 \ , \end{aligned} \tag{3.2.28}$$

implying $n_f = (3/4) n_b$.

A similar computation for the energy density gives

$$\rho_b = \frac{\pi^2}{30} g T^4 \ ; \tag{3.2.29}$$

$$\rho_f = \frac{7}{8}\frac{\pi^2}{30} g T^4 \ . \tag{3.2.30}$$

In the relativistic limit, $E \sim |\mathbf{p}|$; comparing expressions (3.2.21) and (3.2.22), we recover the pressure-density relation for a relativistic gas, as expected:

$$p = \frac{1}{3}\rho \ . \tag{3.2.31}$$

Non-relativistic species

For $T \ll m$, the exponential factor dominates in the denominator of our distribution function, both for bosons and fermions. So, the bosonic and fermionic nature of the particles becomes indistinguishable. Furthermore, we can expand the energy

$$E = m\left(1 + \frac{|\mathbf{p}|^2}{m^2}\right)^{1/2} \cong m + \frac{|\mathbf{p}|^2}{2m} \ , \tag{3.2.32}$$

since $\frac{|\mathbf{p}|^2}{m^2} \ll 1$ in this limit. Now, we can write f as:

$$f(|\mathbf{p}|) \approx \frac{1}{e^{m/T + |\mathbf{p}|^2/2mT}} = e^{-m/T}\frac{1}{e^{|\mathbf{p}|^2/2mT}} \ . \tag{3.2.33}$$

Defining $x = |\mathbf{p}|/\sqrt{2mT}$, we have for the number density:

$$n \cong \frac{g}{2\pi^2} e^{-m/T} (2mT)^{3/2} \int_0^{+\infty} x^2 e^{-x^2} dx \ . \tag{3.2.34}$$

This integral is the definition of the gamma-function,

$$\int_0^{+\infty} x^n e^{-x^2} dx = \frac{1}{2}\Gamma\left(\frac{1+n}{2}\right) \ . \tag{3.2.35}$$

Taking $\Gamma(3/2) = \sqrt{\pi}/2$, we get the final result:

$$n \cong g\left(\frac{mT}{2\pi}\right)^{3/2} e^{-m/T} \ , \tag{3.2.36}$$

which gives the Boltzmann distribution. As expected, massive particles are exponentially rare at low temperatures. At lowest order in the non-relativistic limit, we have $E(p) \approx m$ and the energy density is simply equal to the mass density:

$$\rho \approx mn \ . \tag{3.2.37}$$

To obtain the associated pressure, we note that to leading order $|\mathbf{p}|^2/E \cong |\mathbf{p}|^2/m$; it is then a similar exercise to obtain

$$p = nT \ , \tag{3.2.38}$$

which corresponds to the familiar result for a non-relativistic perfect gas (the Boltzmann constant k_B is hidden in our system of units). Moreover, since $T \ll m$, we have $p \ll \rho$, showing that a non-relativistic gas of particles acts like pressureless dust (i.e. matter).

By comparing the relativistic and the non-relativistic limits, we conclude that the three statistical quantities — number density, energy density and pressure — fall exponentially (with the Boltzmann factor) as the temperature drops below the particle's mass. Again, this means that if equilibrium had persisted until today, any massive particle species would be eventually suppressed and the Universe would be mostly photons.

3.2.1 The Net Particle Number

Before continuing, it is important to restore a finite chemical potential in order to calculate the net particle number; otherwise, the numbers of particles and antiparticles are equal.

Consider, for example, particle-antiparticle annihilation, $X + \bar{X} \leftrightarrow \gamma + \gamma$. We know that the number of photons is not conserved at high temperatures, implying $\mu_\gamma = 0$ (for instance, double Compton scattering, $e^- + \gamma \leftrightarrow e^- + \gamma + \gamma$ is effective at redshifts above $z \cong 2 \times 10^6$ [21]). This means that, if the chemical potential of particle X is μ_X, then the chemical potential of the corresponding antiparticle $\bar{X}$ is

$$\mu_{\bar{X}} = -\mu_X \ . \tag{3.2.39}$$

Let us recalculate the relativistic limit for the particle number density of fermions, with $\mu \neq 0$ and $T \gg m$. We obtain:

$$n - \bar{n} = \frac{g}{2\pi^2} \int_0^\infty p^2 \left(\frac{1}{e^{(p-\mu)/T} + 1} - \frac{1}{e^{(p+\mu)/T} + 1} \right) dp = \frac{gT^3}{6\pi^2} \left[\pi^2 \left(\frac{\mu}{T} \right) + \left(\frac{\mu}{T} \right)^3 \right] \ , \tag{3.2.40}$$

with g denoting the number of intrinsic degrees of freedom of baryons.

Restoring a finite μ in the non-relativistic limit, we get:

$$n = g \left(\frac{mT}{2\pi} \right)^{3/2} e^{-(m-\mu)/T} \ , \tag{3.2.41}$$

so that

$$n - \bar{n} = 2g \left(\frac{mT}{2\pi} \right)^{3/2} e^{-m/T} \sinh \left(\frac{\mu}{T} \right) \ . \tag{3.2.42}$$

3.2.2 Energy and Entropy Density

Let T be the temperature of the photon gas in the early Universe. The total radiation density is the sum over the energy densities of all relativistic species; hence

$$\rho_r = \frac{\pi^2}{30} g_\star(T) T^4 \ , \tag{3.2.43}$$

where $g_\star(T)$ counts the total number of effectively massless degrees of freedom of the plasma:

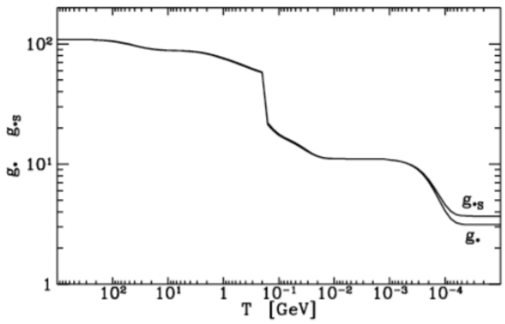


Figure 3.1: Variation of the number of relativistic degrees of freedom, g^* and g^*_S, with temperature according to the Standard Model of Particle Physics [21].

$$g_\star = \sum_{bosons} g_i \left(\frac{T_i}{T}\right)^4 + \frac{7}{8} \sum_{fermions} g_i \left(\frac{T_i}{T}\right)^4 . \tag{3.2.44}$$

Here T_i denotes the effective temperature for any species i. There are two different contributions for this sum: relativistic species in thermal equilibrium with the photons, with $T_i = T \gg m_i$; and relativistic species that are not in thermal equilibrium with the photons (decoupled from the thermal bath), for which $T_i \neq T \gg m_i$. For temperatures $T \gtrsim 10^2$ GeV, all the degrees of freedom of the Standard Model of Particle Physics are in equilibrium, corresponding to $g_\star \approx 107$.

To describe the evolution of the Universe, it is useful to track a conserved quantity: Entropy, as derived in section (3.1.1) from the energy-momentum conservation. To derive an expression for the entropy, we once again resort to the First Law of Thermodynamics (for zero chemical potential):

$$TdS = dE + pdV \ . \tag{3.2.45}$$

Considering $S = S(V, T)$ and $E = E(V, T)$ and using the fact that both the energy and the entropy are extensive quantities (so we can replace the partial derivatives by the quantity itself), we get

$$T\frac{\partial S}{\partial V}dV + T\frac{\partial S}{\partial T}dT - pdV = \frac{\partial E}{\partial V}dV + \frac{\partial E}{\partial T}dT \ \Rightarrow \left(T\frac{S}{V} - p\right) dV + T\frac{\partial S}{\partial T}dT = \frac{E}{V}dV + \frac{\partial E}{\partial T}dT \ . \tag{3.2.46}$$

Equating the terms in dV and dT, this yields:

$$\frac{\partial E}{\partial T} = T\frac{\partial S}{\partial T} \ ; \tag{3.2.47}$$

$$\frac{E}{V} = T\frac{S}{V} - p \ . \tag{3.2.48}$$

From the second equality, we obtain:

$$s = \frac{\rho + p}{T} \ , \tag{3.2.49}$$

defining the entropy density as $s = S/V$. For radiation, with $p_i = \rho_i/3$, this reduces to:

$$s_i = \frac{4}{3}\frac{\rho_i}{T_i} \ . \tag{3.2.50}$$

Summing over all relativistic species, we can use the result (3.2.43) to write the total entropy density of radiation in the early Universe:

$$s = \frac{2\pi^2}{45} g_{\star S}(T) T^3 \ , \tag{3.2.51}$$

where $g_{\star S}(T)$ is the effective number of relativistic degrees of freedom contributing to the entropy. Given that $s_i \propto T_i^3$, this corresponds to

$$g_{\star S} = \sum_{bosons} g_i \left(\frac{T_i}{T}\right)^3 + \frac{7}{8} \sum_{fermions} g_i \left(\frac{T_i}{T}\right)^3 \ . \tag{3.2.52}$$

For species in thermal equilibrium, $g_{\star S} = g_\star$, which is not true for decoupled species. For most of the history of the Universe, however, we can safely use the equality. The difference between these values becomes significant only at low temperatures (see figure 3.1), as neutrinos decouple from the thermal bath.

Entropy conservation implies that $S = a^3 s$ remains constant as the Universe expands, hence

$$g_{\star S} T^3 a^3 = \text{const.} \tag{3.2.53}$$

Away from particle mass thresholds, $g_{\star S}$ is approximately constant and $T \propto a^{-1}$, as expected for relativistic particles ($\rho \propto a^{-4} \propto T^4$). Also, since $s \propto a^{-3}$, we can define the number of particles of a given species in a comoving volume as:

$$N_i \equiv a^3 n_i = \frac{n_i}{s} \ , \tag{3.2.54}$$

where we have redefined the scale factor to absorb constant factors. If no particles are being created or destroyed in that volume, this quantity remains constant.

Using the previous results for the number density of particles in both the relativistic and non-relativistic limits, we can write:

$$N_i = \begin{cases} \frac{45\zeta(3)}{2\pi^4} \frac{g_i}{g_{\star S}} \ , & \text{for} \quad T \gg m_i \ ; \\ \frac{45}{4\sqrt{2\pi}} \frac{g_i}{g_{\star S}} \left(\frac{m_i}{T}\right)^{3/2} e^{-m_i/T} \ , & \text{for} \quad T \ll m_i \ . \end{cases} \tag{3.2.55}$$

For $T \gg m_i$, the conservation of this quantity is obvious: in the primitive thermal soup, all reactions occur in both ways, so the particles which are produced get destroyed almost instantly. In the case of $T \ll m_i$, the number of particles is not fixed until the temperature drops to the value at which reactions which create/destroy them freeze out.

An important consequence of this is that, in the absence of interactions that produce or destroy baryon number, the baryon-to-entropy ratio is conserved:

$$\eta_S = \frac{n_B}{s} = \frac{n_b - n_{\bar{b}}}{s} = \text{const.} \tag{3.2.56}$$

where b and $\bar{b}$ denote baryons and antibaryons, respectively.

Before we discuss the next topic, it is convenient to write the Friedmann equation in the radiation dominated era ($t \lesssim 4 \times 10^{10}$ sec), for which the scale factor evolves as $a(t) \propto t^{1/2}$. Substituting (3.2.43) in (3.1.7), we obtain:

$$H = \frac{\pi}{\sqrt{90}} \sqrt{g_\star} \frac{T^2}{M_P} \ , \tag{3.2.57}$$

where we have defined the reduced Planck mass $M_P = 1/\sqrt{8\pi G} \approx 2.4 \times 10^{18}$ GeV. Using the fact that $H = (1/2t)$, it is a trivial exercise to relate time and temperature:

$$t \approx \left(\frac{T}{\text{MeV}} \right)^{-2} \text{sec} \ . \tag{3.2.58}$$

3.3 The Matter-Antimatter Asymmetry

3.3.1 Evidence for a Baryon Asymmetry

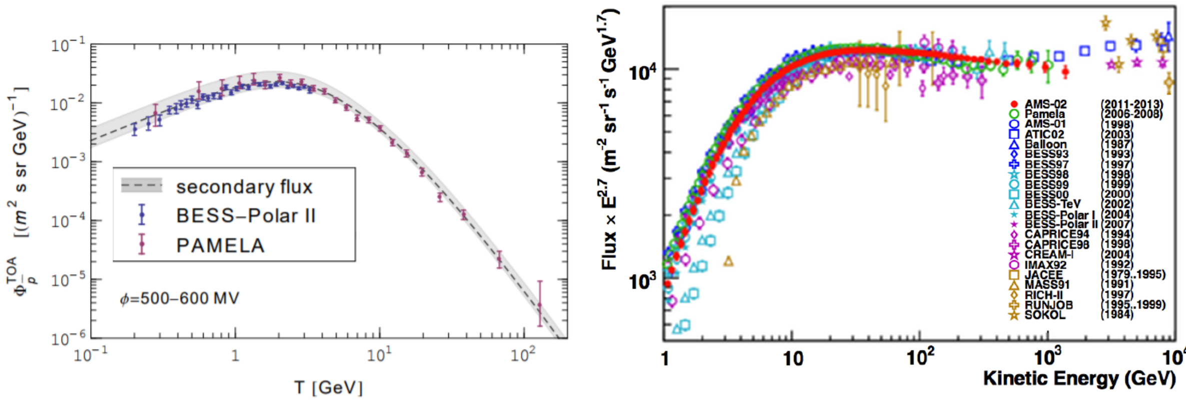
From the discussion of the second chapter, we conclude that CPT invariance ensures the baryon and antibaryon thermal distributions to be equal. This is true, because they have equal masses and the chemical potential can be neglected for most of the thermal history of the Universe (as we will investigate in the next chapters, this is due to the smallness of the baryon-to-photon ratio). This means that if annihilations are efficient in the early Universe (particles and antiparticles quickly annihilate via processes like $p\bar{p} \to \gamma\gamma$) and we start with equal amounts of matter and antimatter (the primitive Universe is neutral), then there would be very few matter in the Universe, which would contain mainly radiation.

The observed Universe, however, is drastically different. We do not observe any bodies of antimatter within the Solar System and, at the same time, cosmological traces of antibaryons are only found in the cosmic ray data. As shown in figure 3.2, the measurements by the Alpha Magnetic Spectrometer (AMS-02), a high energy particle detector designed to study the origin and nature of the cosmic rays up to a few TeV from space — which was installed in the International Space Station in 2011 — and other collaborations yield a flux of antiprotons $\bar{p}$ of about

$$\frac{n_{\bar{p}}}{n_p} \sim 10^{-4} \ . \tag{3.3.59}$$

We might wonder if the dominance of matter over antimatter is only local, with other regions of the Universe having an opposite abundance and giving an overall symmetric Universe. If this were the case, in the boundary between regions with opposite overabundance, annihilations of matter and antimatter would originate an enormous flux that has not been observed (the energetic γ-rays, coming from the decay of $p\bar{p}$ into π-mesons and followed by the decay $\pi^0 \to 2\gamma$).

For these reasons, it is widely accepted that — at least — the observable part of our Universe must have developed an excess of particles over antiparticles, at some point in the cosmological evolution. Moreover, we can conclude from observations that, if domains of matter and antimatter exist in the Universe, they are separated by scales certainly larger than the radius of our galaxy (~ 3 Kpc) and most probably larger than the Virgo cluster (~ 10 Mpc) [24], with sophisticated γ-ray detectors potentially capable of reaching a much more severe bound on the observational scale. A numerical analysis of this



(a) Antiproton flux. (b) The average proton flux, for comparison.

Figure 3.2: Antiproton cosmic ray flux as measured by AMS-02 and other experimental collaborations. The observed antiproton flux is consistent with secondary production due to collisions of baryons, leptons or photons in the interstellar medium (taken from Refs. [22,23]).

problem was performed by Cohen, Rujúla and Glashow (1997), showing that the Universe must consist entirely of matter on all scales up to the Hubble size [25].

General considerations on Nucleosynthesis and the Baryon Number

The baryon number density n_b does not remain constant during the evolution of the Universe; instead, it scales like a^{-3}, where a is the cosmological scale factor[4]. As the photon number density also scales like this (use the result 3.2.53 in equation 3.2.27), one has another good way to define the baryon asymmetry of the Universe, in terms of the quantity

$$\eta \equiv \frac{n_B}{n_\gamma} \,, \tag{3.3.60}$$

where $n_B = n_b - n_{\bar{b}}$ is the difference between the number of baryons and antibaryons per volume and $n_\gamma = 2\frac{\zeta(3)}{\pi^2}T^3$ is the photon number density at some temperature.

The parameter η is essential for determining the present cosmological abundances of light elements (such as H, D, ^{3}He, ^{4}He and ^{7}Li) produced at the Big Bang Nucleosynthesis (BBN) era; and can be used to estimate η_S which has been constant since then, because the Universe is isentropic. This is not quite true for η, which is only constant away from particle mass thresholds (interchanging the entropy density with the photon number, we carry an important g_{*S} factor, which shows variations of about two orders of magnitude). It is important to discuss this in more detail, because several authors use either η or η_S in phenomenological models for the generation of the baryon asymmetry, without explaining with clarity the difference between the two.

It is true that N_B is preserved once baryon-violating interaction freeze out. This is almost true for

[4] We argued that if no particles are created or destroyed in a comoving volume, then $N_i = a^3 n_i$ is constant.

the number of photons, except for some additional ones created during the electron-positron annihilation phase, at $T \approx 1$ MeV, or in the course of stellar evolution. If this variation can be neglected — in the absence of any significant, subsequent annihilations and providing most stellar photons are absorbed by the intergalactic medium — η is approximately constant.

Observational probes strictly constrain the present value of η. This allows us to infer the baryon-to-entropy ratio. The entropy is presently about equally divided between the 2.7 K photons and the three cosmic neutrino backgrounds (with flavours e, μ, τ). Then, the present values for the entropy density and the photon number density are the same up to a constant,

$$\eta_S = \left(\frac{n_\gamma}{s}\right)\eta = \frac{45\zeta(3)}{g_{*S}\pi^4}\eta \approx 10^{-1} \times \eta \ , \qquad (3.3.61)$$

where we have used $g_{*S} \approx 2 + \frac{7\times 6}{8}\left(\frac{T_D^\nu}{T_{CMB}}\right)^4$, $T_D^\nu \sim 2$ MeV being the temperature at which neutrinos decouple from the thermal bath [20]. If we want to map this present value back to the time the baryon-violating forces decoupled from the thermal bath, we should rigorously use η_S.

Observations

The baryon-to-photon ratio is particularly constrained by BBN. The theoretical predictions and experimental measurements are summarized in figure 3.3. The boxes represent the regions which are consistent with experimental measurements, with their associated statistical and systematic uncertainties. The ^{4}He curve is very sensitive to the uncertainty in the neutron lifetime (as discussed in the next section); while the spreads in the curves for D, ^{3}He and ^{7}Li correspond to uncertainties in nuclear cross sections estimated by computational methods [9]. There is acceptable agreement among the abundances when

$$5.8 \times 10^{-10} \leq \eta_{\rm BBN} \leq 6.6 \times 10^{-10}\,(95\%\,\text{C.L.}) \ . \qquad (3.3.62)$$

It might seem uninteresting to worry about numerical values besides the order of magnitude. It is true that the latter suffices to discard statistical fluctuations as the origin of the asymmetry or to quantify the incompatibility with the SM. However, this precision is one of the greatest agreements in Cosmology; also, it turns out that the parameter space of our model using $f(R)$-theories is very sensitive to these numerical factors.

While the η ranges spanned by the boxes in figure 3.3 do not all overlap, they are all within a factor ~ 2 of each other. The lithium abundance corresponds to η values which are inconsistent with that of the deuterium abundance, as well as the less constraining ^{4}He abundance. This discrepancy could simply reflect difficulty in determining the primordial lithium abundance; or it could hint towards new physics, which can be a modification of the cosmological evolution that alters the lithium abundance but not the other mass fractions. Despite this “lithium problem”, the overall concordance is remarkable; the concordant η range (3.3.62) is essentially that implied by the D-mass fraction.

An independent way of measuring the baryon asymmetry is following up the Cosmic Microwave Background. In spite of its 2.7 K uniformity, there are tiny fluctuations (apart from an anisotropy due

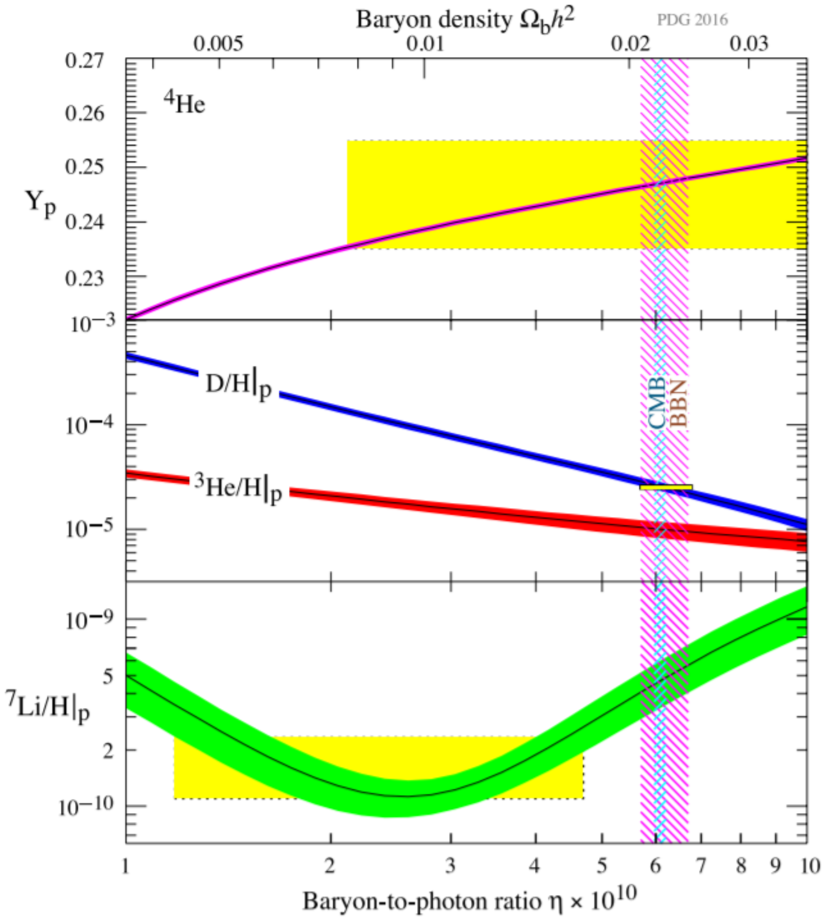


Figure 3.3: The abundances of ^{4}He, D, ^{3}He and ^{7}Li as predicted by the Standard Model of BBN. The bands show the 95% C.L. range. Boxes indicate the observed light element abundances. The narrow vertical band indicates the CMB measure of the cosmic baryon density, while the wider band indicates the BBN concordance range (both at 95% C.L.) [9].

to our motion relative to the CMB frame) within the CMB — of 1 part in 10^5 — that are sensible to the matter density distribution at the time of last scattering; hence, they constrain η. WMAP has been collecting data since 2001 to get more precise measurements of these anisotropies; in particular, the Nine-Year Observations fixed the combination $\Omega_b h^2 = 0.02264 \pm 0.00050$, where h is a conventionally used parameter defined by $h \equiv H_0/100$, $H_0 = 70.0 \pm 2.2$ kms^{-1}Mpc^{-1}. This fixes the baryon-to-photon ratio, $\eta \equiv n_B/n_\gamma$, since the present photon number is calculated for $T = T_{\rm CMB}$ and $n_B = \rho_B/m_H = \Omega_B \rho_C/m_H$, where ρ_B is the baryonic energy density, expressed in terms of the baryonic fraction of the critical density, $\rho_C \equiv (3H_0^2)/8\pi G$. In the end, we get the final constraint from the CMB [26,27]:

$$\eta_{\rm CMB} = (6.19 \pm 0.14) \times 10^{-10} \ . \tag{3.3.63}$$

The concordance between this estimate and the BBN predictions is evident in figure 3.3; these observations are combined from diverse cosmological environments and it is remarkable that two independent tests (the abundances of the light elements and the amplitudes of the acoustic peaks in the CMB angular power spectrum) allow us to compare two measurements of η using very different physics at two widely separated epochs. This allows us to compute the baryon-to-entropy ratio with precision [28]:

$$\eta_S^{\rm obs} \equiv \frac{n_B}{s} \lesssim 9 \times 10^{-11} \ . \tag{3.3.64}$$

In Standard Cosmology, there is no change in the η_S value, after baryon-antibaryon $b, \bar{b}$ annihilations freeze-out, and practically no change in η between BBN and CMB decoupling; thus, the agreement between $\eta_{\rm BBN}$ and $\eta_{\rm CMB}$ is a key test. We have succeeded in the task of extrapolating our laws of physics to an age of ~ 1 sec of the Universe.

3.3.2 The Tragedy of a Symmetric Universe

Although the matter-antimatter asymmetry appears to be very large today, in the sense that $n_B \approx n_b \gg n_{\bar{b}}$, the fact that $n_B/s \approx 10^{-10}$ implies that this asymmetry was once tiny: $n_B \ll n_b$. To see this, we assume for simplicity that nucleons are the fundamental baryons. Earlier than 10^{-6} sec after the BB, temperature was higher than the mass of the nucleon (see equation 3.2.58); thus, $n_N \approx n_{\overline{N}} \approx n_\gamma$. The entropy density is $s \sim g_* n_\gamma \approx 10^2 n_N$. In turn, the constancy of $\eta_S \sim 10^{-10}$ requires $(n_N - n_{\overline{N}})/n_N \approx 10^2 n_B/s \sim \mathcal{O}(10^{-8})$. This means that, during its earliest phase, the Universe was nearly baryon symmetric.

Suppose we start with $\eta = 0$ (or $n_N = n_{\overline{N}}$). We can compute the final number of nucleons that are left over, after annihilations have frozen out. Earlier than 10^{-6} sec after the BB, nucleons and antinucleons were about as abundant as photons. At temperatures $T \lesssim 1$ GeV, the equilibrium abundance (as the temperature falls below m_b but the species are still coupled to the primordial plasma) of nucleons and antinucleons is

$$\frac{n_N}{n_\gamma} = \frac{n_{\bar{N}}}{n_\gamma} \approx \frac{g_N}{2.4} \frac{(2\pi)^{1/2}}{4} \left(\frac{m_N}{T}\right)^{3/2} e^{-m_N/T} \ . \tag{3.3.65}$$

When the Universe cools off, the number of nucleons (antinucleons) decreases as long as the annihilation rate $\Gamma_{\rm ann} \approx n_b \langle\sigma v\rangle_{\rm ann}$ is larger than the expansion of the Universe. The thermally averaged annihilation cross section $\langle\sigma v\rangle_{\rm ann}$ is of the order of $1/m_\pi^2$ with $v \sim 1$ and where m_π is the pion mass. The freeze-out temperature T_F is obtained by setting $\Gamma \approx H$, where the expansion rate of the Universe is given by the Hubble parameter during radiation era, $H \approx 1.66\sqrt{g_*}T^2/m_P$; then, we should solve

$$\begin{aligned} g_N \frac{1}{(2\pi)^{3/2}} x^{1/2} e^{-x} \frac{m_N}{m_\pi^2} &= 1.66 \frac{g_*^{1/2}}{m_P} \\ \Rightarrow x^{-1/2} e^{x} &\sim \frac{8}{3(1.66)} \frac{1}{(2\pi)^{3/2}} \frac{m_N}{m_\pi^2} m_P \approx 6 \times 10^{19} \ , \end{aligned} \tag{3.3.66}$$

where we defined $x = m_N/T$. Solving numerically, this gives $x_F \sim 47(8)$; thus, $T_F \approx 20$ MeV. We used $g_b = 8$ considering the 4 spin states of the nucleon (n, p) system ($\times 2$ taking into account the antiparticles) and $\mathcal{O}(g_*) \sim 10$ for $T \lesssim 1$ GeV[5]. At $T_f \approx 20$ MeV, annihilations thus freeze out, nucleons and antinucleons being so rare that they cannot annihilate any longer. Residual nucleon and antinucleon to photon ratio can be calculated using equation (3.3.65):

$$\frac{n_b}{n_\gamma} \approx (2\pi)^{1/2} x_F^{3/2} e^{-x_F} \sim 10^{-18} \ , \tag{3.3.67}$$

which is much smaller than the value required by Nucleosynthesis (the observed η parameter).

In order to avoid the *annihilation catastrophe*, we may suppose the existence of some mechanism that separated matter from antimatter before $T \approx 32$ MeV, which is the solution of equation (3.3.67) with $\eta \approx 10^{-10}$. This value of temperature corresponds roughly to $t \approx 10^{-3}$ sec. Causality, however, excludes this argument: during the radiation era, the scale factor evolves as $a \propto t^{1/2}$; then we can estimate the size of the region causally connected at this time[6], $d_H^{rad}(t) = 2t$, obtaining a causal region way too small to explain the asymmetry over the galaxy scales. If we take into account inflation, the region of the Universe connected today was connected even at those early times. But, if the processes responsible for the separation of matter from antimatter happened before inflation, then the baryon number would be diluted by an enormous factor; on the contrary, if they happened after, there is no mechanism providing a straightforward way to eliminate the boundaries between matter and antimatter islands.

Attempts to explain the same problem considering statistical (Poisson) fluctuations in the baryon and antibaryon distributions have also failed. To see this, let us estimate how many baryons and photons there are today in the volume that encompasses our galaxy. We consider the sun to be a typical star, with an average density equal to $\rho_\odot \approx 1$ g/cm^3 and volume $V_\odot = 10^{18}$ km^3. Then, there are $\approx (\rho_\odot/m_b)V_\odot \sim 10^{57}$ baryons in a typical star. Since there are $\approx 10^{12}$ stars in a typical galaxy, our galaxy contains $\approx 10^{69}$ baryons; and - using the observed η value - 10^{79} photons. At temperatures $T \gtrsim 1$ GeV,

[5]At about this temperature, $T \sim 200$ MeV, the quarks combine into baryons (protons, neutrons, ...) and mesons (pions, ...); only after this stage of evolution, we can talk about composite particles. Above the QCD phase transition, however, almost all these particles (except the pions) are non-relativistic, reducing $g_* \sim 10$.

[6]The distance to the particle's horizon, which is essentially the maximum distance from which particles could travel to the observer, can be written as $d_H = a(t) \int_0^t \frac{dt'}{a(t')} dt'$.

this comoving volume contained $\approx 10^{79}$ photons, $\approx 10^{79}$ baryons and $\approx 10^{79}$ antibaryons. In order to avoid the annihilation catastrophe, this volume would need an excess of baryons of 10^{69}. Considering the uncertainties in our measurements, we could expect (pure statistical) fluctuations in the number of baryons of the order of the square root of the number of events, that is, $N_b - N_{\bar{b}} \approx \mathcal{O}(\sqrt{N_b}) \approx 10^{40}$: almost 30 orders of magnitude too small!

In conclusion, the Standard Cosmological Model provides no explanation for the smallness of the η parameter, starting from $\eta = 0$. So, there must be some mechanism occurring in the early Universe that produces an overabundance of baryons compared to antibaryons: what we call Baryogenesis. It is a strict requirement that Baryogenesis takes place before Nucleosynthesis, for $T \gtrsim 10$ MeV, so that the initial conditions for the production of light nuclear elements are in place.

In 1967, Sakharov [12] suggested that an initially baryon-symmetric Universe might dynamically develop a baryon excess at some time in its thermal history, after which baryon-antibaryon annihilations would destroy all of the antibaryons, leaving the one baryon per 10^{10} photons that we observe today. In this work, he outlined the three necessary conditions for Baryogenesis: the existence of baryon number (B)-nonconserving interactions; a violation of both C and CP symmetries; and a departure from thermal equilibrium.

3.4 Sakharov Conditions for Baryogenesis

3.4.1 B-Number Violation

It is clear that baryon number conservation must be violated if the Universe begins baryon symmetric and then develops a net B. By the time Sakharov published his work, there was no clear motivation for B-nonconserving processes (for example, we do not see a proton — the lightest baryon — decaying); Grand Unification Theories (GUTs) provide such motivation. Current detectors have been looking for proton decays, such as the Super Kamiokande, which gave a lower bound of approx. 10^{34} years for the mean proton lifetime [29].

The prototype example of a B-violating process is the decay of a heavy boson or a Higgs boson, that is allowed in some extensions of the SM — like GUTs — that unify the strong, weak and electromagnetic interactions. GUT models are based on higher symmetries that are broken at scales $M_{\rm GUT} \sim 10^{16}$ GeV [20], leaving the SM gauge group as the only exact symmetry at lower energies, but justifying the apparent unification of the associated couplings at energies $E > M_{\rm GUT}$. In the process of symmetry breaking (similar to the Higgs mechanism), Higgs and gauge bosons can acquire large masses and decay into quarks and leptons, through processes that violate B, such as:

$$X \to qq \ , \quad X \to \bar{q}\bar{l} \ . \tag{3.4.68}$$

The first decay channel gives a $B = +2/3$ and $L = 0$ final state, while the second yields $B = -1/3$

and $L = -1$. This means that there is not a consistent assignment of a baryon/lepton number to the X boson and that both charges are not conserved in these decays. Note, however, that $B - L = +2/3$ is still a symmetry of the underlying GUT.

Actually, we do not need to invoke GUT models to achieve B-violation. The SM conserves B classically, but there is a global quantum anomaly of the chiral current under which B-conservation can be violated. Though it has not been observed experimentally, baryon-number violating processes are plausible in the Standard Model: they emerge as a consequence from the inclusion of instanton fields in the SM Lagrangian. This is a non-perturbative effect that gives interesting outcomes for Particle Physics. In the presence of an instanton, the vacuum of the theory becomes an infinitely degenerate state and the different subspaces (which are not homotopic equivalents) are separated by energy barriers. So, even though they cannot be surpassed in classical theory, the system can move to a different vacuum through a quantum tunneling process. This process is heavily-suppressed at low energies (the decay of the proton would require nearly the age of the Universe to happen), but it is feasible at earlier points in the Universe history.

The chiral anomaly, in the presence of an instanton field, allows processes with:

$$\begin{aligned} &\Delta E = \Delta M = 1 \ ; \\ &\Delta u + \Delta d^C = 3 \ ; \\ &\Delta u' + \Delta s^C = 3 \ ; \end{aligned} \tag{3.4.69}$$

where E is the electronic number, M is the muonic number and u, u', d^C and s^C are the two generation SM quarks (Cabibbo rotated). These conclusions can be inferred from t' Hooft's 1976 work, *Symmetry Breaking through Bell-Jackiw Anomalies* [30]. For example, in the presence of instantons, an allowed decay is:

$$p + n^C \to e^+ + \bar{\nu}_\mu \ , \tag{3.4.70}$$

which violates both B and L. The probability of such decay, in this kind of models, goes with the square of the instanton transition amplitude between two vacuum states, which is of the order of $e^{-16\pi^2/g}$ [31], where $g = e/\sin(\theta_W)$ is the weak coupling constant, written in terms of the electromagnetic coupling and the weak mixing angle (of the electroweak theory). This gives deuteron a lifetime of about 10^{225} sec $\approx 10^{218}$ years! These enormous numbers are characteristic of models with instantons.

3.4.2 C- and CP-Violation

Simple baryon number violation is not enough to explain matter-antimatter asymmetry, if charge conjugation C is a symmetry of the interactions. Let us consider a generic B-number violating reaction $X_0 \to Y_0 + Z_B$, with X_0 representing a generic initial state with vanishing baryon number, Y_0 denoting all particle states in the final state also with a vanishing baryonic charge and Z_B corresponding to all produced particles with an overall B. If C is conserved, we expect the rate of this process to be the same as for the C-conjugated process, involving the antiparticles:

$$\Gamma[X_0 \to Y_0 + Z_B] = \Gamma[\bar{X}_0 \to \bar{Y}_0 + \bar{Z}_B] \ . \tag{3.4.71}$$

This means that a baryon excess B is being produced at the same rate as the opposite excess $-B$, resulting in no net baryon asymmetry. So, C-violation is a necessary condition, although not a sufficient one.

Consider a hypothetical B-number violating scenario where our boson X can decay into either left-handed or right-handed quarks, $X \to q_L q_L$ and $X \to q_R q_R$. Under C, $q_L \to \bar{q}_L$, while under CP we have $q_L \to \bar{q}_R$. Similarly, $q_R \to \bar{q}_R$ under C and $q_R \to \bar{q}_L$ under CP. In this case, violation of C ensures that

$$\Gamma[X \to q_L q_L] \neq \Gamma[\bar{X} \to \bar{q_L}\bar{q_L}] \ , \tag{3.4.72}$$

but if CP is a symmetry of Nature, we have

$$\Gamma[X \to q_L q_L] + \Gamma[X \to q_R q_R] = \Gamma[\bar{X} \to \bar{q_L}\bar{q_L}] + \Gamma[\bar{X} \to \bar{q_R}\bar{q_R}] \ . \tag{3.4.73}$$

Then, if we have equal numbers of X particles and $\bar{X}$ antiparticles initially, we will end up with the same number of quarks and antiquarks, even though an asymmetry between left-handed and right-handed particles may be produced. So, CP needs to be violated as well.

CP violation has been observed in $K^0 - \bar{K}^0$ system, for example; however, a fundamental understanding of the origin of CP violation (and how that amount could be "fed up" to match the observed asymmetry) is still lacking; hopefully additional studies on Baryogenesis can shed some light on the subject [20].

3.4.3 Departure from Thermal Equilibrium

The departure from thermal equilibrium is essential for a non-vanishing baryon asymmetry, because the equilibrium average of B vanishes:

$$\begin{aligned} \langle B \rangle_T &= \mathrm{Tr}\left(e^{-\beta H} B\right) = \mathrm{Tr}\left[(CPT)(CPT)^{-1} e^{-\beta H} B\right] \\ &= \mathrm{Tr}\left[e^{-\beta H}(CPT)^{-1} B (CPT)\right] = -\mathrm{Tr}\left(e^{-\beta H} B\right) \\ &= -\langle B \rangle_T \ , \end{aligned} \tag{3.4.74}$$

where $\beta = 1/T$. We have used the facts that the Hamiltonian H commutes with CPT, while B is odd under CPT (odd under C, even under P and T). Hence, $\langle B \rangle_T = 0$ in thermal equilibrium. This means that B violating processes must be out-of-equilibrium in the Universe, in order to establish a baryon asymmetry dynamically.

3.5 Big Bang Nucleosynthesis

3.5.1 Equilibrium Mass Fractions

As a first step to understanding Big Bang Nucleosynthesis, we consider the consequences of nuclear statistical equilibrium among the light nuclear species. Note that, in spite of heavy elements being generated in the interior of stars via nuclear fusion reactions, the lighter elements such as deuterium D, ^{3}He, ^{4}He and ^{7}Li cannot be produced in this way (the corresponding ratios approach zero in young stars). In fact, stellar Nucleosynthesis occurs subsequently, being responsible for the formation of heavier elements, due to the fusion of hydrogen and helium by the process of evolution of the star core composition. The products of Stellar Nucleosynthesis are released to the interstellar medium through mass loss events, like the planetary nebulae phase of low-mass stars or supernovae explosions. Returning to our discussion, this only means that the observed abundances of light elements had to be present in the primordial gas.

As usual, in kinetic equilibrium, the number density of a nonrelativistic nuclear species $A(Z)$ with Z protons and $A-Z$ neutrons, A being the mass number, is given by:

$$n_A = g_A \left(\frac{m_A T}{2\pi}\right)^{3/2} \exp\left(\frac{\mu_A - m_A}{T}\right) , \tag{3.5.75}$$

where μ_A is the chemical potential. If the nuclear reactions that produce nucleus A from Z protons and $A-Z$ neutrons occur rapidly compared to the expansion rate, chemical equilibrium is also obtained, hence

$$\mu_A = Z\mu_p + (A-Z)\mu_n . \tag{3.5.76}$$

We use it to express the exponential factor in (3.5.75) in terms of the proton and neutron densities:

$$\begin{aligned} \exp\left(\mu_A/T\right) &= \exp\left[\left(Z\mu_p + (A-Z)\mu_n\right)/T\right] \\ &= n_p^Z n_n^{A-Z} \left(\frac{2\pi}{m_N T}\right)^{3A/2} 2^{-A} \exp\left[(Zm_p + (A-Z)m_n)/T\right] . \end{aligned} \tag{3.5.77}$$

Here we set $m_P \approx m_n \approx m_N$ in the pre-exponential factors, keeping the exact masses in the exponentials. The proton and neutron degrees of freedom are considered to be the spin ones, given by $(2s+1)$; both particles have spin $s = 1/2$, which gives $g = 2$.

Recalling the binding energy of a nucleus is defined as

$$B_A = Zm_p + (A-Z)m_n - m_A , \tag{3.5.78}$$

and substituting (3.5.77) in (3.5.75), we get for the abundance of species A:

$$n_A = g_A A^{3/2} 2^{-A} \left(\frac{2\pi}{m_N T}\right)^{3(A-1)/2} n_p^Z n_n^{A-Z} \exp\left(B_A/T\right) , \tag{3.5.79}$$

with the approximation $m_A \approx A m_N$.

Since particle number densities in the expanding Universe decrease as a^{-3} (for constant number per comoving volume), it is useful to define the mass fraction of each nuclear species as:

$$X_A \equiv \frac{n_A A}{n_N} \quad \text{with} \quad n_N = n_p + n_n + \sum_i A_i n_{A_i} \ , \tag{3.5.80}$$

using the total nucleon density n_N and satisfying

$$\sum_i X_i = 1 \ . \tag{3.5.81}$$

Using this definition and noting that $n_p^Z n_n^{A-Z}/n_N = X_p^Z X_n^{A-Z} n_N^{A-1}$, we can write the mass fraction of species A as:

$$X_A = g_A[\zeta(3)^{A-1}\pi^{(1-A)/2}2^{(3A-5)/2}]A^{5/2}\,(T/m_N)^{3(A-1)/2}\,\eta^{A-1}X_p^Z X_n^{A-Z}\exp\left(B_A/T\right) \ , \tag{3.5.82}$$

where $\eta = n_N/n_\gamma$ is the usual baryon-to-photon ratio (all baryons are either in the form of free nucleons or in bound states).

The fact that $\eta \ll 1$ (very high entropy per baryon) is of serious importance to primordial Nucleosynthesis. It means that nuclei with $A > 1$ are much less abundant and that Nucleosynthesis takes place later than we would naively expect considering only the binding energies of nuclear species.

3.5.2 Initial Conditions

The ratio of neutrons to protons is particular interesting in what concerns the outcome of primordial Nucleosynthesis, as essentially all the neutrons available become incorporated into ^{4}He (at $T \approx 0.1$ MeV). At higher temperatures, $T \gg 1$ MeV, protons and neutrons are kept in thermal equilibrium by the weak interaction:

$$\begin{aligned} &n \longleftrightarrow p + e^- + \bar{\nu}_e \ ; \\ &\nu_e + n \longleftrightarrow p + e^- \ ; \\ &e^+ + n \longleftrightarrow p + \bar{\nu}_e \ . \end{aligned} \tag{3.5.83}$$

When the rates of these interactions are higher than the Hubble parameter H, chemical equilibrium is obtained,

$$\mu_n + \mu_\nu = \mu_p + \mu_e \ , \tag{3.5.84}$$

from which follows

$$n_n/n_p \equiv n/p = \exp\left[-Q/T + (\mu_e - \mu_{\nu_e})/T\right] \ , \tag{3.5.85}$$

where $Q \equiv m_n - m_p \approx 1.293$ MeV.

We now need some estimation for the chemical potentials of electrons and electron neutrinos. In the relativistic limit, note that

$$n_{e^-} - n_{e^+} = \frac{2T^3}{6\pi^2}\left[\pi^2\left(\frac{\mu_e}{T}\right) + \left(\frac{\mu_e}{T}\right)^3\right] \ . \tag{3.5.86}$$

At the same time, based upon the charge neutrality of the Universe, the number of electrons (or $n_{e^-} - n_{e^+}$) equals the number of protons:

$$n_{e^-} - n_{e^+} = n_p \approx n_B = \eta \cdot n_\gamma = \eta \frac{2}{\pi^2}\zeta(3)T^3 \ . \tag{3.5.87}$$

Since η is very small, $\mu \ll T$, so we can drop the cubic term in (3.5.86) to get

$$\frac{\mu_e}{T} \approx \frac{6}{\pi^2}\zeta(3)\eta \ . \tag{3.5.88}$$

Thus, we conclude that $\mu_e/T \sim \eta \sim 10^{-10}$.

The electron neutrino number density is similarly related to μ_ν/T; however, no relic neutrino background has been detected. We will assume analogously that $\mu_{\nu_e} \ll T$, so that the difference in the number of neutrinos and antineutrinos is small. Then, the equilibrium value of the neutron-to-proton ratio reads

$$\left(\frac{n}{p}\right)_{\text{eq}} = e^{-Q/T} \ , \tag{3.5.89}$$

at sufficient high temperatures.

3.5.3 Weak Interaction Rate

Now consider the rates for the weak reactions that interconvert neutrons and protons; the rates for these reactions are found by integrating the square of the nuclear matrix element for a given process, weighted by the available phase space densities of particles, while preserving four-momentum conservation and using the usual Lorentz invariant volume element. For example, the rate (per nucleon) for the process $pe \to \nu n$ is given by

$$\begin{aligned}\Gamma_{pe\to\nu n} = \int & f_e(E_e)[1 - f_\nu(E_\nu)]|\mathcal{M}|^2_{pe\to\nu n} \cdot (2\pi)^4\delta^4(p_p + p_e - p_\nu - p_n) \\ & \times \frac{d^3p_e}{2E_e(2\pi)^3}\frac{d^3p_\nu}{2E_\nu(2\pi)^3}\frac{d^3p_n}{2E_n(2\pi)^3} \ .\end{aligned} \tag{3.5.90}$$

All the processes in (3.5.83) have in common the same factor from the nuclear matrix element for the β-decay of the neutron,

$$|\mathcal{M}|^2 \propto G_F^2(1 + 3g_A^2) \ , \tag{3.5.91}$$

where G_F is the Fermi constant and $g_A \approx 1.27$ is the axial-vector coupling of the nucleon, a correction factor describing its internal structure. This factor can be expressed in terms of the mean neutron lifetime, which — after a lengthy calculation performed in appendix A — is found to be:

$$\tau_n^{-1} \equiv \Gamma_{n\to pe\nu} = \frac{G_F^2}{2\pi^3}(1 + 3g_A^2)m_e^5\lambda_0 \ , \tag{3.5.92}$$

where

$$\lambda_0 \equiv \int_1^q d\xi\xi\sqrt{\xi^2 - 1}(\xi - q)^2 \approx 1.636 \tag{3.5.93}$$

is simply a numerical factor coming from the phase space integral for neutron decay.

It is true that all β-decay processes involve the same species, so they have in common the same factor from the nuclear matrix; however, considering a particular reaction rate, as equation (3.5.90) for the $pe \to \nu n$ decay, we need to take into account how many electrons and neutrinos are available at a given time (or temperature) of the Universe evolution. In this case, because we are considering the "$\to$" reaction, we need to quantify how many electrons are available to react and how many neutrinos can be destroyed; this weight is given by the distribution functions in equation (3.5.90). Using the same dimensionless quantities defined in appendix A, $q = Q/m_e$ and $\xi = E_e/m_e$, and designating two more, $z = m_e/T$ and $z_\nu = m_e/T_\nu$, we can write the product $f_e[1 - f_\nu]$ as:

$$f_e[1 - f_\nu] = \frac{1}{1 + e^{\frac{E_e}{T}}} \times \frac{1}{1 + e^{-\frac{E_\nu}{T_\nu}}} = \frac{1}{\left[1 + e^{\xi z}\right]\left[1 + e^{(q-\xi)z_\nu}\right]} \ , \tag{3.5.94}$$

where we have used energy conservation to rewrite the energy of the neutrino.

The total decay rate then follows as:

$$\Gamma_{pe\to\nu n} = (\tau_n \lambda_0)^{-1} \int_q^\infty d\xi \frac{\xi(\xi^2 - 1)^{1/2}(\xi - q)^2}{[1 + e^{\xi z}][1 + e^{(q-\xi)z_\nu}]} \ . \tag{3.5.95}$$

In particular, at sufficient high temperatures ($T \gg Q, m_e$), at which neutrinos have not yet decoupled from the thermal bath ($T_\nu \sim T$), the integral above is approximately equal to

$$\int_0^\infty d\xi \frac{\xi^4}{[1 + e^{\xi z}][1 + e^{-\xi z}]} = \frac{1}{z^5} \int_0^\infty dy \frac{y^4}{[1 + e^{y}][1 + e^{-y}]} = (T/m_e)^5 \, \frac{7}{30}\pi^4 \ , \tag{3.5.96}$$

where, in the intermediate step, we defined $y = \xi z$.

Hence, in the high-temperature limit,

$$\Gamma_{pe\to\nu n} \to \frac{7}{60}\pi(1 + 3g_A^2)G_F^2 T^5 \ . \tag{3.5.97}$$

By comparing Γ to the expansion rate of the Universe during the radiation era, $H \approx 1.66 g_*^{1/2} T^2/m_p$, we find that

$$\Gamma/H \approx (T/R)^3 \ , \tag{3.5.98}$$

where R represents the combined constants entering this ratio,

$$R \equiv \left[\frac{7\pi/60(1 + 3(1.26)^2)(1.166 \cdot 10^{-5})^2 \times (1.22 \cdot 10^{19})}{1.66\sqrt{10.75}}\right]^{-1/3} \text{GeV} \approx 10^{-3} \text{ GeV} \ . \tag{3.5.99}$$

We conclude finally that

$$\Gamma/H \sim (T/\text{ MeV})^3 \ , \tag{3.5.100}$$

for $T \gtrsim m_e$. Thus, at temperatures greater than the freeze-out temperature, $T_F \sim 1$ MeV, one expects the neutron-to-proton ratio to be equal to its equilibrium value, which at temperatures much greater than 1 MeV ($n/p \sim \exp[-Q/T] \to 1$) implies $X_n \sim X_p$.

3.5.4 Formation of the Primordial Elements

At temperatures of about $T \gtrsim 10$ MeV, the energy and the number density were dominated by relativistic and therefore effectively massless particles: electrons, positrons, neutrinos and photons. At this early time, all particles are kept in thermal equilibrium via weak interactions, by their rapid collisions. All weak rates are much larger than the expansion rate, so the neutron-to-proton ratio (equation 3.5.89) is very close to unity and $T_\nu = T$. At this early epoch, the baryon density is too low: the light elements are in statistical equilibrium, but they have very small abundances, due to the fact that η is so small. Also, as first argued by Gamow (1946), the conditions necessary for rapid nuclear reactions existed only for a very short period of time [32].

The idea behind the origin of the primordial elements was then that heavier nuclei have to be built sequentially from lighter nuclei in two-particle reactions involving neutrons and protons, so that deuterium is formed first; once deuterons are available, helium nuclei can be formed and so on. The fact that the abundances of the light elements did not begin to build up until temperatures of much less than $T \sim 1$ MeV is, to some extent, explained recurring to the mitigating "deuterium bottleneck" process: as the temperature drops, the equilibrium abundances rise fast, becoming large later for nuclei with small binding energies, which is the case of the deuteron, the first element to be formed directly from neutrons and protons. Because this happens, heavier nuclei with larger binding energies, whose equilibrium abundances would become large earlier, "have to wait" to be formed.

Consider the following system of neutrons, protons, deuterons, ^{3}He nuclei, ^{4}He nuclei and ^{12}C nuclei. For each of these species, equation (3.5.82) reads:

$$X_n/X_p = \exp\left(-Q/T\right) \; ; \tag{3.5.101}$$

$$X_2 = 16.3(T/m_n)^{3/2}\eta \exp\left(B_2/T\right) X_n X_p \; ; \tag{3.5.102}$$

$$X_3 = 57.4(T/m_n)^3\eta^2 \exp\left(B_3/T\right) X_n X_p^2 \; ; \tag{3.5.103}$$

$$X_4 = 113(T/m_n)^{9/2}\eta^3 \exp\left(B_4/T\right) X_n^2 X_p^2 \; ; \tag{3.5.104}$$

$$X_{12} = 3.22 \times 10^5 (T/m_n)^{33/2}\eta^{11} \exp\left(B_{12}/T\right) X_n^6 X_p^6 \; ; \tag{3.5.105}$$

$$1 = X_n + X_p + X_2 + X_3 + X_4 + X_{12} \; . \tag{3.5.106}$$

For purposes of illustration, consider also their binding energies (table 3.1):

A_Z	B_A	g_A
^{2}H	2.2 MeV	3
^{3}H	6.92 MeV	2
^{3}He	7.72 MeV	2
^{4}He	28.3 MeV	1
^{12}C	92.2 MeV	1

Table 3.1: The binding energies of some light nuclei [21].

The evolution of the mass fractions for this system is displayed in figure 3.4. Note that the equilibrium abundances of ^{4}He and ^{12}C (nuclei with large binding energies) are very small until temperatures that are less than 0.3 MeV. This is due to the high entropy of the Universe: the formation of these elements is suppressed by large powers of $\eta \ll 1$.

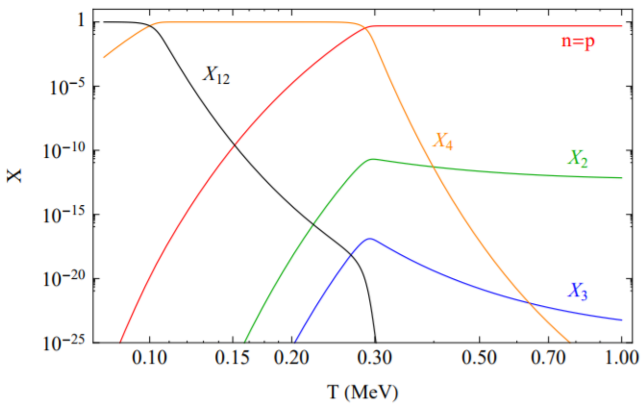


Figure 3.4: Mass fractions in thermal equilibrium for a system of neutrons, protons, D, ^{3}He, ^{4}He and ^{12}C as a function of the temperature (using the simplification $X_n \sim X_p$) [20].

Although for temperatures of a few MeV, composite nuclei are favored on energetic grounds (the average binding energy per nucleon varies between 1 and 8 MeV), entropy considerations favor free nucleons; consequently, the temperature needs to drop somewhat below 1 MeV for these abundances to increase.

Figure 3.4 also hints that in thermal equilibrium heavier elements will eventually become the dominant form of baryonic matter. If this were the case, we would end up with very little Hydrogen and Helium around, in contrast with the observations we have today of the Universe. Once again, departures from thermal equilibrium are the key to understand the cosmological evolution.

3.5.5 Estimate of the Helium Abundance

Weak interactions freeze-out around $T = T_F \sim 1$ MeV, leading to neutrino decoupling from the plasma and later on to annihilation of electrons and positrons (via electromagnetic processes) that transfer their entropy to photons, raising the photon temperature relative to that of the neutrinos. The weak interactions that convert neutrons into protons and vice-versa also freeze-out, yielding:

$$\left(\frac{n}{p}\right)_{t=t_F} \sim e^{-Q/T_F} \approx \frac{1}{6} \ . \tag{3.5.107}$$

As the Universe cools downs, neutrons can still decay, $n_n = n_n(T_F)e^{-t/\tau_n}$, where $\tau_n \approx 886$ sec. This will decrease the neutron-to-proton ratio until the time Nucleosynthesis actually takes place, for $T_{NUC} \approx 0.1$ MeV, or equivalently $t_{NUC} \approx 100$ sec, yielding:

$$\left(\frac{n}{p}\right)_{t=t_{NUC}} \sim \frac{1}{6}e^{-t_{NUC}/\tau_n} \approx \frac{1}{7} \ . \tag{3.5.108}$$

If ^{4}He was to track its equilibrium abundance, then X_4 would approach unity at a temperature of about 0.3 MeV. This assumes, however, that the synthesis of ^{4}He proceeds through a chain of reactions which are also in thermal equilibrium; mainly:

$$\begin{aligned} \mathrm{D} + \mathrm{D} &\rightarrow {}^3\mathrm{He} + \mathrm{n} \ , \\ {}^3\mathrm{He} + \mathrm{D} &\rightarrow {}^4\mathrm{He} + \mathrm{p} \ ; \end{aligned} \tag{3.5.109}$$

$$\begin{aligned} \mathrm{D} + \mathrm{D} &\rightarrow {}^3\mathrm{H} + \mathrm{p} \ , \\ {}^3\mathrm{H} + \mathrm{D} &\rightarrow {}^4\mathrm{He} + \mathrm{n} \ ; \end{aligned} \tag{3.5.110}$$

and

$$\mathrm{D} + \mathrm{D} \rightarrow {}^4\mathrm{He} + \gamma \ . \tag{3.5.111}$$

These are not fast enough to keep up with the rapidly increasing "equilibrium demand" for ^{4}He.

The nuclear reaction rates, which are proportional to $n_A \langle \sigma |v| \rangle$, cannot keep up with expansion for two reasons: **(1)** The rate at which ^{4}He is produced depends on the number densities of the lighter elements involved in these reactions, D, ^{3}He, and ^{3}H , which in thermal equilibrium are very small. From figure 3.4, X_i is between 10^{-20} and 10^{-12} at this stage. So, the number densities of these fuels, $n_A = (X_A/A)\eta n_\gamma$ are small too, contributing to a decreasing decay rate. This is the "deuterium bottleneck" at work. **(2)** Overcoming the Coulomb-barrier introduces an exponential suppression of the decay rate which is approximately given by [21]:

$$\langle \sigma v \rangle \propto \exp\left[-2\bar{A}^{1/3}(Z_1 Z_2)^{2/3} T_{\mathrm{MeV}}^{-1/3}\right] \ , \tag{3.5.112}$$

where $\bar{A} = A_1 A_2/(A_1 + A_2)$. This must be considered, as one needs two protons to tunnel through this repulsive barrier to get close enough for the nuclear strong force to take over and form a nucleus of ^{4}He.

This effect is less important for temperatures $T \lesssim 0.1$ MeV. For these reasons, around $T \sim 0.5$ MeV, the amount of ^{4}He drops below its equilibrium value.

On the contrary, since deuterium is formed directly from neutrons and protons, it can follow its equilibrium abundance as long as enough free neutrons are available. However, since the deuterium binding energy is rather small, the deuterium abundance becomes large rather late. Only when there is enough deuterium, can helium be produced. To get a rough estimate of this temperature, we can solve equation (3.5.102) for the temperature when $X_2 \sim 1$, which yields $T_{NUC} \sim 0.1$ MeV. Then, surpassing the "bottleneck problem" and since the binding energy of helium is larger than that of deuterium, helium production is favored and essentially all available neutrons are quickly bound into ^{4}He. The resulting mass fraction is easy to estimate:

$$X_4 = \frac{4n_4}{n_N} \approx \frac{4(n_n/2)}{n_n + n_p} = \frac{2(n_n/n_p)}{1+(n_n/n_p)} \ . \tag{3.5.113}$$

Using equation (3.5.108), this gives:

$$X_4 \approx \frac{2/7}{1+1/7} = \frac{1}{4} = 0.25 \ , \tag{3.5.114}$$

which is in remarkable agreement with the observational value [9]:

$$X_4 = 0.249 \pm 0.009 \ . \tag{3.5.115}$$

Accurate results are estimated by computing the time at which the actual nuclear reaction rates freeze out and the species decouple, yielding $X_2 \sim X_3 \sim 10^{-5}$ and $X_7 \sim 10^{-10}$ for ^{7}Li, which is also produced. These are the results of computational code when $\eta_{10} \equiv (\eta/10^{-10})$ is in the range of $1 - 10$ [9]. In fact, following up the chain reactions that produce ^{4}He, we can still have other strong reactions occurring, such as [33]:

$$\begin{aligned}
&{}^4\mathrm{He} + {}^3\mathrm{He} \rightarrow {}^7\mathrm{Be} + \gamma \ ; \\
&{}^4\mathrm{He} + {}^3\mathrm{H} \rightarrow {}^7\mathrm{Li} + \gamma \ ; \\
&{}^7\mathrm{Be} + \mathrm{n} \rightarrow {}^7\mathrm{Li} + \mathrm{p} \ ; \\
&{}^7\mathrm{Li} + \mathrm{p} \rightarrow {}^4\mathrm{He} + {}^4\mathrm{He} \ .
\end{aligned} \tag{3.5.116}$$

The reaction chain then proceeds (from the already analysed synthesis of ^{4}He) along stable and long-lived isotopes — compared to the Nucleosynthesis timescale, which corresponds to a few minutes — towards large mass numbers; the increasing Coulomb-barrier suppression, however, prevents the formation of heavier nuclei. Furthermore, the mass numbers $A = 5$ and $A = 8$ form “bottlenecks” — they have no stable isotopes, so the lack of sufficient densities of these elements hinders the production of heavier ones. The $A = 5$ bottleneck is crossed with the reactions $^4\mathrm{He} + {}^3\mathrm{He}$ and $^4\mathrm{He} + {}^3\mathrm{H}$, which form a small fraction of ^{7}Be and ^{7}Li. Their abundances remain so small that the reactions crossing the $A = 8$ bottleneck (like $^7\mathrm{Be} + {}^4\mathrm{He} \rightarrow {}^{11}\mathrm{C} + \gamma$ and $^7\mathrm{Li} + {}^4\mathrm{He} \rightarrow {}^{11}\mathrm{B}$) can be safely ignored. Thus BBN produces essentially ^{2}H, ^{3}H, ^{3}He, ^{4}He, ^{7}Li and ^{7}Be. Of these, ^{3}H and ^{7}Be are unstable so they decay after Nucleosynthesis into

^{3}He and ^{7}Li (e.g., ^{7}Be becomes ^{7}Li via electron capture). Substantial amounts of both D and ^{3}He are left unburnt ($X_2, X_3 \sim 10^{-5}$), as the reactions that burn them to ^{4}He freeze out.

In the end, comparing to present abundances, BBN has produced cosmologically significant quantities of ^{2}H, ^{3}He, ^{4}He and ^{7}Li.

In stellar cores, the triple-α reaction can occur, using ^{4}He nuclei to produce carbon-12; this can then react to produce heavier elements, such as N, O and even Fe, in stars with a large mass, via the CNO cycle, that end up enriching the interstellar medium through the explosion of a supernova, for example. So, the ejected remains of stellar Nucleosynthesis not only alter the light element abundances of primordial elements, but also produce these heavy metals. Thus, in order to measure light element abundances which are closer to primordial, one seeks astrophysical locals with low metal abundances [9]: D is revealed in the spectrum of low-metallicity quasar absorption systems; ^{4}He is observed in clouds of ionized hydrogen (called H-II regions), etc.

3.5.6 Sensitivity to Cosmological Parameters

The successful predictions of BBN pose tight constraints on the cosmological parameters and thus on extensions of the Standard Model that may change them. We discuss the sensitivity of the light element abundances to three important parameters:

1. **Neutron half-life**

   From previous calculations, we verified that all relevant weak interaction rates go with $\Gamma \sim T^5/\tau_n$. These rates determine the temperature at which neutrons and protons decouple, which is given by the condition $\Gamma \sim H \propto T^2$, which implies $T_F \propto \tau_n^{1/3}$; then, a larger τ_n corresponds to an earlier freeze-out and a larger value of $(n/p)_{T_F}$, increasing the prediction for X_4.

2. **Number of relativistic species**

   Since $H \propto g_*^{1/2} T^2$, an increase in the input value of the number of relativistic species also leads to an earlier freeze out of the neutron-to-proton ratio, since $T_F \propto g_*^{1/6}$. Hence, more ^{4}He would be produced. Moreover, in the SM one expects[7] $g_* = 5.5 + (7/4)N_\nu$, where N_ν is the number of neutrino species. The current bounds on X_4 give $N_\nu = 3.13 \pm 0.31$ [9], in agreement with the existence of only three SM neutrinos; this shows how the dependence of T_F upon g_* can be used to study the possible existence of additional light particle species and how Cosmology and Particle Physics can be so intrinsically connected.

3. **Baryon-to-photon ratio**

   Finally, let us discuss the important dependence of the primordial abundances on the baryon-to-photon ratio, as $X_A \propto \eta^{A-1}$. This means that for larger values of η the abundances of D, ^{3}He, ^{3}H

[7] Use equation (3.2.44) considering $\gamma, e^\pm$ and ν to be the only massless species.

build up earlier (X_A gets slightly closer to unity), and thus ^{4}He synthesis starts earlier, when the neutron-to-proton ratio is larger, resulting in more ^{4}He. However, around the time ^{4}He synthesis begins in earnest, for $T = T_{NUC} \approx 0.1$ MeV, the neutron-to-proton ratio is only slowly decreasing, due to neutron decays, and thus the sensitivity of ^{4}He production to η is only slight (much smaller than the dependence on the previous two parameters). But the amount of D and ^{3}He left unburnt depends strongly on η, decreasing with increasing η: the nuclear rates that convert these elements into ^{4}He, $\Gamma \propto X_{2,3}(\eta n_\gamma) \langle\sigma|v|\rangle$, become small as X_2 and X_3 become small until they freeze-out; a large η delays the freeze-out, favoring the conversion of these elements into helium-4. This analysis is in agreement with the observations, as confirmed by figure 3.3. BBN ends up constraining $\eta = (5 - 7) \times 10^{-10}$.

Chapter 4

Spontaneous Baryogenesis

Let us go back to section 3.4, where we discussed the three necessary conditions for Baryogenesis.

Although B violation and C-CP violation have been investigated only within the context of Particle Physics models, the third condition — departure from thermal equilibrium — can be discussed in a more general way.

Following the lead of Sakharov, many authors have tried to explain the value of the asymmetry in terms of out-of-equilibrium decay scenarios, such as models of GUT Baryogenesis. Cohen and Kaplan showed that this is avoidable [34]: the baryon asymmetry can actually occur while baryon violating interactions are still in thermal equilibrium[1], through a general mechanism of *spontaneous baryogenesis* based on the classical motion of a scalar field along its potential. The key to this scenario is a dynamical, temporary violation of CPT. Naively, we would not expect an isotropic and homogeneous Universe to distinguish between "left" and "right", have a preferred direction, etc. In the early Universe, however, these symmetries might have not been already established.

4.1 Basic Setup

The crucial feature of this model is adding a term to the Lagrangian density of the form

$$\mathcal{L}_{\text{eff}} = \frac{1}{\Lambda}\partial_\mu \phi j_B^\mu \ , \tag{4.1.1}$$

where j_B^μ is the net baryon number current, ϕ is a scalar field, and Λ is an energy scale. This corresponds to a dimension-five effective operator[2] (irrelevant in four dimensions), assumed to result from some unspecified dynamics at a scale $T \sim \Lambda$.

[1] This means that we know how to — easily — compute the asymmetry parameter, using our standard thermodynamic quantities (number and entropy densities). However, we still need a subsequent departure from thermal equilibrium: otherwise, the baryon asymmetry produced in equilibrium will eventually disappear.

[2] This results from simple dimensional analysis: $[\phi] = M$, $[\partial] = M$ and $[j^\mu] = [\phi\partial\phi^*] = M^3$.

This term, when non-zero, will dynamically violate CPT invariance. The argument is as follows. To begin with, we consider ϕ to be spatially constant, to assure the homogeneity of space is not broken; then we have

$$\frac{1}{\Lambda}\partial_\mu \phi j^\mu_B = \frac{1}{\Lambda}\dot{\phi}(n_b - n_{\bar{b}}) \ . \tag{4.1.2}$$

It is easy to see that this new term shifts the energy of a baryon relative to that of an antibaryon; it has the effect of giving baryons a chemical potential $\mu_b = -\dot{\phi}/\Lambda$ and antibaryons a chemical potential $\mu_{\bar{b}} = -\mu_b = \dot{\phi}/\Lambda$:

$$\mathcal{L}_{\text{eff}} = \mu n_B \ . \tag{4.1.3}$$

Thus, in thermal equilibrium, there will be a non-zero net number density, as given by equation (3.2.40):

$$n_B = n_b - n_{\bar{b}} \approx \frac{g_b \mu_b T^2}{6} \sim -\frac{\dot{\phi} T^2}{\Lambda} \ , \tag{4.1.4}$$

which yields

$$\frac{n_B}{s} \sim -\frac{\dot{\phi}}{g_\star \Lambda T} \ , \tag{4.1.5}$$

so that the asymmetry parameter is fixed by the behavior of the field ϕ and the choices for the mass scale and the temperature at which B-violating interactions freeze out.

In this scenario, ϕ is initially displaced from its equilibrium point, where the first derivative of its potential vanishes; as it evolves toward that point, $\dot{\phi} \neq 0$ which leads to a non-negligible chemical potential for baryons. Eventually, when ϕ settles into the minimum of its potential, η becomes zero. However, η will only track its equilibrium value as long as B-violating interactions are occurring rapidly ($\Gamma \gtrsim H$). If these interactions become ineffective before $\dot{\phi} = 0$, a non-zero value of η will freeze out, leaving the Universe with a permanent residual baryon asymmetry.

4.2 CPT-violating Interaction

As noted before, a homogeneous and isotropic Universe should not distinguish between C, P and T transformations. However, these symmetries might have not been already settled in the early times: after all, an expanding Universe at finite temperature might violate both Lorentz invariance and time reversal.

Vector fields that gain a vacuum expectation value (recall section 2.6.1) are not contained in the SM and there is no observational evidence for such a field at a present time, but they can arise in approaches to more fundamental physics, like in string theory, and there is no obvious reason to avoid them in the first stages of our Universe's evolution. The VEV of a vector field selects a preferred direction in the vacuum, thus violating Lorentz symmetry.

The time reversal violation can be understood from the simple fact that, in the first seconds, processes can be distinguishable from the ones occurring backwards, because the Universe has not achieved an isotropic state yet.

It is possible for this lack of symmetry to lead to *effective* CPT violating interactions among the baryons. This can only happen in an evolving Universe and only for a finite time; here, the Hubble parameter has a significant role, determining the size of the CPT violations. At $T \to 0$, all interactions must be CPT invariant if derivable from a relativistic field theory which has a Lorentz invariant ground state (if we stick with the interpretation that CPT violation arises due to a dynamical breakdown of Lorentz symmetry).

Then, the baryon asymmetry is generated due to the effective coupling (4.1.1). If we assume the existence of some neutral scalar field, it is natural that it acquires effective couplings to other fields, in particular to the baryon current, unless it is forbidden by some symmetry. Cohen and Kaplan refer to the ϕ particle as the *thermion*. The thermion field can be given cause to develop a slowly varying time derivative as the Universe cools $\dot{\phi}/\Lambda \equiv \mu$, which can be treated as a classical background. Of course the baryon number must still be violated; otherwise, this shift in the energy spectra would not allow a baryon asymmetry to develop, since the total B-charge could not change. However, if there are baryon violating interactions in thermal equilibrium, interaction (4.1.3) will cause the baryons and antibaryons to equilibrate with different thermal distributions[3].

We already established that the CPT breaking term assigns for each particle/antiparticle pair an extra energy $\Delta E \equiv \dot{\phi}/\Lambda$ per particle and $-\Delta E$ per antiparticle. While in equilibrium, this can be alternatively interpreted as particles obtaining a chemical potential μ and antiparticles $-\mu$. This is valid, because — in the equilibrium particle phase distribution function —, an energy shift ΔE is equivalent to assigning a $\Delta\mu$. The "classical background" attribute of this effective potential, though, must be confirmed by the fact that $\mu \sim \dot{\phi}$ behaves (nearly) as a constant.

The true ground state of the Universe should be Lorentz invariant and unchanged by CPT transformation, with $\langle \partial_\mu \phi \rangle = 0$. Thus, μ must approach zero as the Universe cools. In contrast, B-violating forces must drop out-of-equilibrium before μ relaxes to zero, so that the net baryon asymmetry $n_B \sim \mu T^2$ becomes frozen in place and we are left with a non-zero net baryon asymmetry.

Next, we discuss how the thermion can develop an expectation value for its velocity. This can be attained using some symmetry breaking mechanism. One can consider a scalar field which develops a vacuum expectation value. If a continuous symmetry is broken, a Goldstone boson is produced, which will be at some initial value θ_i generally differing from the true minimum of the potential. It will then slowly roll to its true minimum.

[3]This would not be the case if CPT was, at these early times, a symmetry of Nature. The equilibrium particle phase distributions are given by $f(p) = [\exp(\mu/T + E/T) \pm 1]^{-1}$. In equilibrium, processes like $b + \bar{b} \leftrightarrow \gamma + \gamma$ imply $\mu_{\bar{b}} = -\mu_b$ and the fact that $n_B \approx T^3(\mu/T) = \eta n_\gamma$ gives $\mu/T \approx 0$. Since $E^2 = p^2 + m^2$ and $m_b = m_{\bar{b}}$ by the CPT Theorem, it follows that, in thermal equilibrium, $n_b \equiv n_{\bar{b}}$ (recalling that the number density is given by $n = \int d^3p f(p)/(2\pi)^3$). Thus, with CPT invariance, baryons and antibaryons have equal thermal distributions.

4.3 Dynamics of ϕ during Baryogenesis

To realize this mechanism, we consider, following up Cohen and Kaplan's work, that the thermion arises from the spontaneous breakdown — at a temperature $T_0 = f$ — of some approximate U(1) symmetry; and we rename the resultant Goldstone boson as the thermion field.

Then, we can write the Lagrangian density of the thermion in terms of an angular variable $\theta = \phi/f$. In order to obtain an expectation value for the time-derivative of the Goldstone field, a potential term is also included:

$$\mathcal{L}_\theta = \frac{1}{2} f^2 \partial_\mu \theta \partial^\mu \theta - V(\theta) - q \partial_\mu \theta j^\mu_B \ , \tag{4.3.6}$$

where q is a dimensionless constant and $V(\theta)$ is some arbitrary potential that can be approximated by

$$V(\theta) \approx \frac{1}{2} m^2 f^2 \theta^2 \ , \quad m \ll f \ . \tag{4.3.7}$$

The argument leading to this procedure is as follows: we start with an unspecified action for a scalar field that respects U(1) invariance, such as $S = \int d^4x \mathcal{L}$ with $\mathcal{L} = \frac{1}{2}(\partial_\mu \phi^*)(\partial^\mu \phi) + m^2 \phi^* \phi +$ other terms in $U(\phi, \phi^*)$. Imagine the minimum of the potential U is at $\langle \phi \rangle = f e^{i\theta}$. Then, as usual, we can expand the action around the VEV: $\phi \to [f + r(x,t)]e^{i\theta}$. The resulting Lagrangian density becomes $\mathcal{L}' = \frac{f^2}{2}(\partial_\mu \theta)(\partial^\mu \theta) + m^2 f^2 +$etc. (other dynamical field arises, but we are interested in the θ-Lagrangian), in which the last term does not correspond to a mass, but is only a constant with no physical meaning. As the effect of a U(1) transformation on the original field $\phi \to e^{i\epsilon}\phi$ translates into a phase shift $\theta \to \theta + \epsilon$ for the thermion, the vacuum state $|0\rangle$ is no longer preserved. The resulting action can include other terms, arising from the expansion of the bare Lagrangian near the minimum, such as $\partial_\mu \theta j^\mu_B$ [35]. The fact that we also include a quadratic term $V(\theta)$ (explicitly breaking the U(1) symmetry) has nothing to do with the mechanism of spontaneous symmetry breaking, which explains a possible origin for the thermion field: it is introduced, so that the desirable dynamics are obtained, in the development of the action functional.

We assume that at a scale $T_0 = f$, the coupling corresponding to the last term in equation (4.3.6) is the most relevant to the Universe dynamics; operators of higher dimensions can be suppressed by powers of f. It is also assumed that the theory has baryon violating interactions which go out of thermal equilibrium at a temperature T_D, during the evolution of the Universe. It is not necessary, for the present model, to know the form of these interactions, nor the specific content of the theory at high energies.

The equation of motion can be obtained from the Lagrangian for a spatially constant θ field in a FRW Universe, taking the usual derivatives as covariant ones. The final result is:

$$\ddot{\theta} + 3H\dot{\theta} + m^2\theta = \frac{q}{f^2}(\dot{n_B} + 3Hn_B) \ , \tag{4.3.8}$$

with the Hubble parameter given by

$$H = \frac{1}{2t} = \frac{\kappa}{M_{Pl}} T^2 \ , \tag{4.3.9}$$

where we have defined $\kappa \equiv 1.66\sqrt{g_*} \approx 17$ at this stage. Notice that, as previously noted, without the linear term in θ (equation 4.3.8) we would not obtain a damping. The importance of this effect is discussed along this section.

4.3.1 Symmetry Breaking and Thermal Baryon Number

We want to discuss the thermodynamics of this model as the Universe cools down from the symmetry breaking temperature, f. In order to interpret the interaction term as an effective chemical potential that shifts the baryon and antibaryon energy levels,

$$\mu = -q\dot{\theta} \ , \tag{4.3.10}$$

we need to assure that the rate of change $\dot{\theta}$ is sufficiently slow; that is, the typical scale of baryon violating interactions must be fast enough to maintain thermal equilibrium:

$$\tau_{\Delta B}(T) < \mu_B/\dot{\mu_B} = \dot{\theta}/\ddot{\theta} \ . \tag{4.3.11}$$

From statistical mechanics, we have already obtained the baryon density, which to first order in μ/T is given by[4]:

$$n_B = -\frac{g}{6}q\dot{\theta}T^2 \equiv -\frac{1}{12}Bq\dot{\theta}T^2 \ , \tag{4.3.12}$$

where B is defined as the absolute value of the baryon number summed over each spin degree of freedom. Then, the equation of motion (4.3.8) can be rewritten as

$$\ddot{\theta} + 3H\dot{\theta} + m^2\theta = -\frac{1}{12}q^2B[(\partial_t + 3H)(\dot{\theta}H/H_0)] \ , \tag{4.3.13}$$

where H_0 is the initial value of the Hubble parameter at $T_0 = f$. Because of the factor H/H_0, the r.h.s. is apparently only significant at early times, so we can neglect it to study the thermodynamic regime (after a few instants of evolution). Changing the variable to $z \equiv mt = m/2H \propto m/T^2$, equation (4.3.13) becomes:

$$\theta''(z) + \frac{3}{2z}\theta'(z) + \theta(z) = 0 \ , \tag{4.3.14}$$

defining $\theta' \equiv \frac{d\theta}{dz} = \frac{1}{m}\frac{d\theta}{dt}$ and using $H\frac{d\theta}{dt} = \frac{1}{2t}\frac{d\theta}{dt} = \frac{m}{2z}(m\theta')$.

Hence, the solution of the equation (4.3.8), in the limit $q \to 0$, is given by

$$\theta(z) = z^{-1/4}[A\boldsymbol{J}_{1/4}(z) + B\boldsymbol{Y}_{1/4}(z)] \ , \tag{4.3.15}$$

where $\boldsymbol{J}_\alpha$ and $\boldsymbol{Y}_\alpha$ are Bessel functions of the first and second kind, respectively; A and B are constants to determine from the proper initial conditions. We impose $\theta = \theta_0$, $\dot{\theta} = 0$ at $z = z_0 \equiv m/2H_0 \ll 1$, meaning[5] that the thermion's mass is small enough so that its motion is strongly Hubble damped at the

---

[4] In this thermal regime, we will find $\mu \approx m \ll T$, so this is a very good approximation.

[5] Cohen and Kaplan actually show that the damping term (in equation 4.3.13) has only a negligible effect on the evolution of the field for all times, provided that $z_0 \lesssim 0.1$ [34].

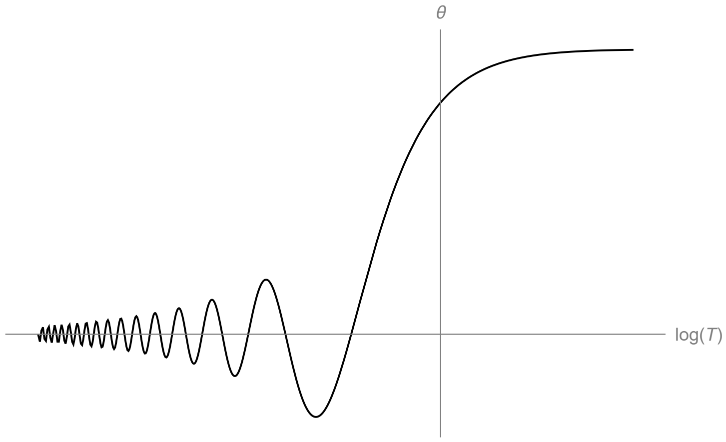


Figure 4.1: The behaviour of θ versus the logarithm of the temperature, using equation (4.3.15) with $B \ll A$, as imposed by the initial conditions.

initial temperature, f. Then we find, to leading order in z_0:

$$\begin{aligned} A &= \theta_0 2^{1/4}\Gamma(5/4) \ , \\ B &= \theta_0 \frac{\pi}{5 \times 2^{1/4}\Gamma(5/4)} z_0^{5/2} \ll A \ . \end{aligned} \tag{4.3.16}$$

This solution for θ is graphed in figure 4.1. During the period when $H \gg m$, the field changes slowly; eventually, H becomes comparable with m and the θ field begins oscillating.

As the Universe cools below $T_0 = f$, the baryon density remains in thermal equilibrium with an effective potential μ_B, which is decreasing. Eventually, we reach the decoupling temperature T_D, when the baryon violating interactions drop out of equilibrium. From then on, the baryon distribution function will still look thermal, with a constant μ_B fixed at the time of the decoupling. To calculate the baryon-to-photon ratio, we just need to evaluate the effective chemical potential at $z_D \equiv m/2H_D$.

The decoupling temperature is determined by equating the baryon violating interaction time to the expansion rate,

$$\tau_{\Delta B}(T_D) \approx H_D^{-1} \ . \tag{4.3.17}$$

Moreover, to guarantee the distributions remain thermal, we must also ensure that the condition for slow variation of the effective chemical potential (4.3.11) is always satisfied. From the definition of $z \equiv mt \propto m/H$, we verify that

$$\frac{\dot{\mu_B}}{\mu_B} = \frac{\ddot{\theta}}{\dot{\theta}} \sim \begin{cases} H \ , & \text{at early times} \\ m \ , & \text{at somewhat later times} \end{cases} \ . \tag{4.3.18}$$

Recall, however, that this treatment is only valid before the field enters its oscillating phase, around $H \approx m$. The system above can then be translated into the condition[6]

$$\tau_{\Delta B}(T) < \min(H^{-1}, m^{-1}) \ , \tag{4.3.19}$$

[6]When the θ field starts its oscillating phase, μ as given by equation (4.3.10) can no longer be interpreted as an effective

for slow variation of μ. This, in conjunction with equation (4.3.17), guarantees that the decoupling occurs before θ begins to oscillate; or equivalently, that $z_D < 1$. Consequently, we can approximate the Bessel functions in equation (4.3.15) by the first terms of their series expansion:

$$\theta'(z) = \frac{\theta_0}{25} z \left[-10 + 10 \left(\frac{z_0}{z}\right)^{5/2} + 10 \left(\frac{z_0}{z}\right)^{11/2} z_0^2 - \frac{\sqrt{2}\pi}{\Gamma(5/4)^2} z_0^{5/2} \right] . \tag{4.3.20}$$

Noting that $z \geq z_0$, while $z_0 \ll 1$, the two last terms can be neglected; then, we can write:

$$\dot{\theta} = -m \frac{2}{5} \theta_0 z \left[1 - \left(\frac{z_0}{z}\right)^{5/2} \right] , \tag{4.3.21}$$

returning to our initial variable $\dot{\theta} = m\theta'(z)$.

The final result for the baryon-to-photon ratio is thus:

$$\eta = \frac{\pi^2}{60\zeta(3)} \left(\frac{\theta_0 q B}{2\kappa}\right) \frac{m^2 M_{Pl}}{T_D^3} \left[1 - \left(\frac{T_D}{f}\right)^5 \right] , \tag{4.3.22}$$

writing z in terms of the temperature, $z = m/2H = mM_{Pl}/2\kappa T^2$. Note that η depends only weakly on the symmetry breaking scale, f; it is specified primarily by the free parameters of the model, m and T_D, which are constrained by the condition $z_D < 1$. Observational probes constrain $\eta \lesssim 10^{-9}$; hence, equation (4.3.22) gives a rough estimate for the decoupling temperature, placing T_D above 10^8 GeV.

4.3.2 Oscillating Baryon Number

In this section, we consider the subsequent development of the baryon asymmetry, after baryon violating interactions have fallen out of equilibrium. Below T_D, the interaction of the thermion with the baryon current can no longer be interpreted as an effective potential for baryon number ($\dot{\theta}$ changes significantly in this period); so, the equations of motion must be solved directly. To do that, we first consider the Lagrangian (6.2.11) without the expansion of the Universe, as suggested in the original work.

The interaction term can be integrated by parts, resulting in

$$q\theta\partial_\mu j_B^\mu . \tag{4.3.23}$$

It is clear that if the baryon current was conserved, the divergence would vanish; however, since the baryon number conservation is assumed to be broken, the divergence might be replaced by the operator that violates baryon number. Instead of specifying this operator, we assume that it gives rise to a decay of the thermion field[7], with a width Γ. Then, we can approximate the effect of the decay of the motion of

chemical potential (it changes too rapidly and its average drops to zero); hence, the baryon violating forces must drop out of equilibrium before, when $H \gg m$ ($z \ll 1$) or, in the limiting case, $H \approx m$ ($z \lesssim 1$). In this way, at the decoupling temperature, μ (or $\dot{\phi}$) — which is still "well-behaved" — is fixed and fixes the baryon asymmetry.

[7] Note that this is exactly the same idea of the reheating mechanism that needs to occur after inflation, so that the Universe can undergo the standard cosmological phases of its evolution.

the thermion field (due to its baryon-violating interactions) by including an extra term in the equations of motion:

$$\ddot{\theta} + m^2\theta + \Gamma\dot{\theta} = 0 \ . \tag{4.3.24}$$

This is the usual equation of motion for a damped oscillator: θ oscillates with a frequency m with an amplitude that decreases exponentially with time.

We have described the argument that has been used in the literature to explain the dynamics of θ during its oscillatory phase. To sum up: we need to solve the equation of motion for the thermion, $\ddot{\theta} + m^2\theta = -\partial_\mu j^\mu_B f^{-2}$, which — since it is oscillating — will produce both baryons and antibaryons with different number densities (the current is not conserved). To calculate the asymmetry in this case, it is assumed that the former equation of motion for θ with the back reaction of the produced particles is given by (4.3.24), which correctly describes the decrease of the amplitude of motion due to the production of baryons/antibaryons.

Comparing the last equation with (4.3.8), we can relate the evolution in time of the baryon density with the decay width of the field:

$$\dot{n_B} = -\Gamma f^2 \dot{\theta} q^{-1} \ . \tag{4.3.25}$$

Integrating, we get

$$n_B(t) = n_B(t_0) - \Gamma f^2 q^{-1}[\theta(t) - \theta(t_0)] \ . \tag{4.3.26}$$

One immediate remark is that, as the θ field oscillates, so does the baryon number.

Another conclusion is that, even though θ damps out to zero as $t \to \infty$, the net baryon density left is not zero, but instead depends on how far from the true minimum of the potential (assumed to be at $\theta = 0$) we started the θ field.

This result becomes clearer if we consider the graph of $\dot{\theta}$, in figure 4.2. Note that θ decays at $H \sim \Gamma$, so — to see it oscillate — we represent the regime where $H \sim m \gtrsim \Gamma$.

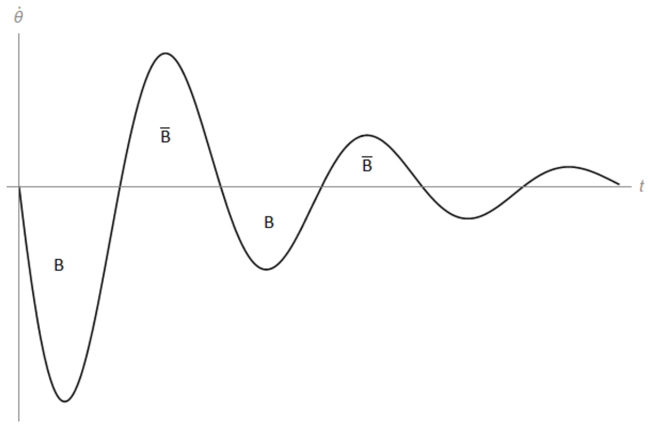


Figure 4.2: $\dot{\theta}$ versus time, below decoupling. During this period, the field can also produce an asymmetry: $\dot{\theta} < 0$ implies baryon production, while $\dot{\theta} > 0$ implies antibaryon production.

When the velocity of the field is negative, the decay of the thermion produces mostly baryons; when the velocity is positive, the decay produces mostly antibaryons (under the assumption that the thermion can only decay into one of these species). Equation (4.3.26) indicates that, starting from an initial $n_B(t_0)$, we are either decreasing or increasing the baryon number, depending on the sign of $\dot{\theta}$. As n_B corresponds to the difference between the number of baryons and antibaryons at some time, we end up producing mostly one or the other.

Another important implication of this analysis is that the regions where the velocity is positive are not equivalent to the regions where the velocity is negative (the amplitude of the θ field is being damped). Hence, as $t \to \infty$, the baryon density is determined by the net asymmetry, which is in turn only a function of the difference in initial and final positions of θ.

The final step is to include the effects of the expanding Universe. This can be done by replacing any volume factors by the corresponding comoving ones, proportional to a^3; the effect must be the same as simply including the appropriate metric factors in the Lagrangian density and considering the fact that — when the Hubble parameter falls below m — we can neglect the Hubble friction. Then, the expansion serves only to dilute the amount we have previously calculated by the ratio of the scale factor cubed. Using our previous definition, $N_B = (n_B(t_0) + \Gamma f^2 \Delta\theta q^{-1})V$, where $\Delta\theta \equiv \theta(t_0) - \theta(\infty)$ and V refers to the volume, the net particle density becomes

$$n_B = \frac{N_B}{\left(\frac{a}{a_0}\right)^3 V} = \left(\frac{a_0}{a}\right)^3 \left[n_B(t_0) + \Gamma f^2 \Delta\theta q^{-1}\right] , \tag{4.3.27}$$

concerning the period from the oscillating starting mode until large times, when $\theta(t)$ approaches the true minimum. From the baryon density, we want to extract the baryon-to-photon ratio: as usual, the photon number density is given by

$$n_\gamma = \frac{2\zeta(3)}{\pi^2} T^3 , \tag{4.3.28}$$

the expression for the baryon-to-photon ratio thus reads

$$\eta = \frac{\pi^2}{2\zeta(3)T^3}\left[n_B(t_0) + \Gamma f^2 \Delta\theta q^{-1}\right] \times \left(\frac{a_0 T_0}{aT}\right)^3 \left(\frac{T}{T_0}\right)^3 . \tag{4.3.29}$$

Since the evolution of the Universe is approximately isentropic[8], being the total entropy equal to $S = g_\star a^3 T^3$, we may treat the factor aT as a constant.

Since the oscillating phase of the θ field begins after the baryon violating interactions fall out of equilibrium[9], $T(t_0)$ is just $T_D \lesssim f$. Hence, the final result is:

$$\eta = \eta_0 + \frac{q^{-1}\Gamma f^2 \Delta\theta}{T_D^3} \frac{\pi^2}{2\zeta(3)} . \tag{4.3.30}$$

[8]Although the entropy contained in the θ field is being transferred after the decay, we can argue that, in the regime of temperatures we are dealing with, this fraction can be neglected.

[9]During the thermodynamic regime, the velocity of the field acts as the slope along which baryon violating forces interact; after the correct amount of asymmetry is fixed, the field must decay to recover the true CPT-invariant ground state of the Universe, with $\langle\dot{\theta}\rangle = 0$.

In this regime, all baryon violating interactions are no longer able to affect the result; the only important baryon violating effect is the conversion of the energy stored in the θ field oscillations into baryons.

Thus, we conclude that, provided decoupling happens before the thermion has rolled past its true minimum, the effect of the subsequent evolution of the thermion is to increase the baryon-to-photon ratio. As the baryon asymmetry produced in this regime just adds to the result of the previous section, we also conclude that the generation of an asymmetry from oscillations may occur independently of thermodynamic generation; that is, we can still produce a baryon asymmetry, even if $\eta_0 = 0$.

4.4 Problems with Homogeneity

Let us review some features of this model.

Basically, we studied the process that allows the baryon asymmetry to be generated by the classical motion of a scalar field in its potential. In this model, we found two distinct regimes:

- the **thermodynamic regime**, corresponding to small accelerations of the thermion field ($\ddot{\theta}/\dot{\theta} \ll \tau_{\Delta B}^{-1}$);
- the **oscillatory regime**, corresponding to relative large accelerations of the field ($\ddot{\theta}/\dot{\theta} \gg \tau_{\Delta B}^{-1}$).

In both cases, whether baryons or antibaryons are produced is an accident of the initial conditions. This brings arbitrariness to our discussion: the field can travel clockwise or counter clockwise to get from its initial value to its final position; either possibility is equally likely. This problem arises from the fact that the model has no explicit CP violation.

With this in mind, it might seem that the mechanism is useless: the initial conditions for the θ field cannot possibly be homogeneous over distance scales greater than the horizon size at the initial temperature[10], $T_0 \approx f$. This means we could go to one region of space where more baryons had been produced and then to another region having mostly antibaryons. Averaging all these contributions, we would end up with $\langle \theta_0 \rangle \approx 0$, or certainly a value too small to account for the amount of η we need to match observations [34].

This difficulty is easily solved by including CP violation in the original Lagrangian. For example, we could include CP violating self-interactions for the thermion. This would destroy the $\theta \to -\theta$ symmetry of the potential $V(\theta)$. It might seem that the introduction of CP violation would have been inevitable, following the arguments of Sakharov. However, in this symmetry breaking model, this is only necessary to produce a net baryon asymmetry when averaged over space.

The need to include explicitly CP violation can be, however, avoidable, if we resort to inflation. By

[10] We can only argue that θ_0 is the same over distance scales of the order of $d_H \sim 2t_0$, t_0 referring to the initial temperature $T_0 \gtrsim T_D$ ($\gtrsim 10^8$ GeV). But this gives $t_0 \lesssim 10^{-22}$ sec, corresponding to a ridiculous value for the horizon size!

introducing a period of exponential expansion into the thermal history of the Universe, our entire horizon today can be fashioned out of a single correlated region at early times.

Essentially, with inflation, we allow random regions with different initial conditions for the θ field to come in thermal equilibrium before going out of causal contact. For this to happen, we need to assume that the symmetry breaking scale f (at which the thermion is produced) is above the reheating temperature following inflation; then, the visible Universe today could be spawned from a region that had a homogeneous initial value θ_0. We also need to require the value of θ at the end of inflation to be $\mathcal{O}(1)$. This is necessary, because any baryons produced before or during inflation will be washed out by the rapid expansion of the Universe. Recalling the previous analysis, this constraint is satisfied if the velocity of the field is still slowly varying by the time inflation is terminating. Since this happens in the regime where $m \ll H$, this is equivalent to demanding the mass of the θ field to be much smaller than the Hubble parameter during the inflationary phase.

It is important to note that one cannot achieve sufficient baryon number simply by assuming that our present causally connected Universe evolved from an inflated region of space, that had a slightly excess of baryons. After the dominance of the baryon violating forces, the baryon number is conserved; consequently, the baryon density in such a scenario would be diluted by the ratio of initial to final volumes, an enormous amount. The reason why we can still have an adequate asymmetry with this model, while requiring inflation to solve the problem of the initial conditions, is that, with these constraints, the field can enter its oscillating phase after inflation. This allows the production of a non-vanishing η — with the energy stored in the thermion oscillations — even if the asymmetry produced during the thermodynamic regime η_0 is almost entirely diluted.

One might ask if the baryon asymmetry could be produced by the inflaton itself. If the inflaton is identified with θ, then thermodynamic baryogenesis is out of question: that only occurs during a slow rolling period which, for the inflaton, corresponds to the period of exponential expansion; thus, the resultant asymmetry would be exponentially diluted. It is conceivable, however, that the inflaton could generate the baryon number while it oscillates in its potential. Then the discussed model applies, but the scale parameter f will be related to inflation.

Chapter 5

Modified Gravity

In this chapter, we introduce modified gravity models to our discussion, as the former is required to enforce a mechanism for gravitational baryogenesis — as will be made transparent in the following discussion.

We present $f(R)$-theories and then include a non-minimal coupling between curvature and matter. The NMC introduces the matter Lagrangian in the gravitational field equations, alters the thermodynamic conservation law associated with the energy-momentum tensor and enlarges the degrees of freedom in the cosmological background, enabling a possible energy exchange between the two.

5.1 The Need For Modified Gravity

It is now more than a hundred years since Einstein formulated his General Theory of Relativity (GR). This has changed profoundly our understanding of the Universe and is the framework for the development of our Standard Model of Cosmology. GR revolutionized our idea of spacetime, changing it from a fixed plane where dynamics is played out to a structure in which matter and geometry are deeply linked. Although it is a brilliant theory, raised upon first principles, with accurate predictions (like the recently detected gravitational waves) and by far the one with most experimental support, recent observations suggest that GR is not a finished theory.

Contemporary Cosmology is faced with the outstanding challenge of understanding the existence and nature of the so-called dark components of the Universe — dark energy and dark matter — which are thought to be dominating the present Universe, although they were never directly detected. Dark energy is required to explain the accelerated cosmological expansion, accounting for about 70% of the energy content of the Universe; dark matter corresponds to approximately 27% of the total energy, and its existence can justify, for instance, the flattening of galactic rotation curves and cluster dynamics.

The Standard ΛCDM Model was built to accommodate the observational data, consisting of a Universe which evolves according to GR, but with an additional dark energy component, represented by a cosmological constant (Λ) with a negative equation of state $p = -\rho$, and Cold Dark Matter (CDM), a non-baryonic type of matter that interacts very weakly with baryonic matter. This model includes also an inflationary scenario to account for the early exponential growth of the Universe, which is needed to explain the large-scale structure formation, among other fundamental problems, like the uniformity of the CMB radiation, the flatness of the Universe and the dissolution of primordial magnetic monopoles [20].

Instead of including these extra components in the Universe, we could understand them as hints towards an incompleteness of the fundamental laws: this is the motivation for modified gravity models. Some proposals posit extensions of the Friedmann equation that include higher order terms in the energy density. Another straightforward approach lies in considering changes on the fundamental action functional, replacing the linear scalar curvature term in the Einstein-Hilbert action by a function of the scalar curvature, $f(R)$. Due to the arbitrariness in the choice of this function, a wide range of interesting phenomena might be produced. It is possible to explain, for example, the accelerated expansion of the Universe and structure formation without adding those unknown dark components.

In spite of the phenomenological nature of this work, the successes of these theories in explaining the current mysteries of the Universe are impressive. For example, the model with $f(R) = R + \alpha R^2$ ($\alpha > 0$) was proposed by Starobinsky (1980) to explain inflation. The αR^2 term dominates the dynamics in the early Universe (where the curvature is large) and leads to inflation, but is negligible at late times, where we recover GR. On the contrary, the model with $f(R) = R - \alpha/R^n$ ($\alpha, n > 0$) has been extensively studied to explain the late-time acceleration of the Universe; whilst it is faced with several instabilities, many authors have been attempting to derive conditions for the cosmological viability of $f(R)$ dark energy models [36].

From the cosmological point of view, the idea behind these theories is that, instead of the scalar curvature in the gravity action, the $f(R)$ function admits an expansion as a Taylor series, with each term dominating the dynamics at a specific scale. In local regions whose densities are much larger than the homogeneous cosmological density, this generalized function needs to be close to GR for consistency with local constraints; for example, Solar System tests could restrict the possible forms of these $f(R)$ theories.

In the attempt to generalize the model, one can include not only a non-linear scalar curvature term in the Einstein-Hilbert Lagrangian density, but also a non-minimal coupling (NMC) between the scalar curvature and the matter Lagrangian density $\mathcal{L}$. Note that matter and geometry are only implicitly related in the Einstein-Hilbert action, since the covariant terms in $\mathcal{L}$ are constructed by contraction with the metric, like $g^{\mu\nu}\chi_{,\mu}\chi_{,\nu}$ for the kinetic term of some scalar matter field. In regions where the curvature is high, which in GR ($R \sim H^2 \sim \rho$) corresponds to high energy density or pressure, the implications of such model could again deviate considerably from those predicted by Einstein's theory. Isolating the effect of the NMC, considered to be some power-law, it was shown that the accelerated phase of the Universe might be reproduced for suitable choices of the exponent, and dark matter can be mimicked[1].

[1]Although current evidence, such as the Bullet Cluster, support the existence of dark matter.

Furthermore, in the presence of this NMC, the same equation of state arises both for dark matter and dark energy, suggesting some level of unification between the two [37,38].

5.2 $f(R)$ Theories

5.2.1 The Model

To begin with, we should start by considering the action that automatically generalizes GR. Thus, instead of the linear curvature term minimally coupled to the metric, $f(R)$ theories consider a generalized function of the scalar curvature R:

$$S = \int [\kappa f(R) + \mathcal{L}]\sqrt{-g}d^4x \ , \tag{5.2.1}$$

where $\kappa = c^4/16\pi G$, $\mathcal{L}$ denotes the Lagrangian density of matter and g is the metric determinant.

Varying the action with respect to the metric, we obtain the modified field equations

$$2\kappa F(R)G_{\mu\nu} = T_{\mu\nu} + 2\kappa\Delta_{\mu\nu}F(R) + \kappa f(R)g_{\mu\nu} - \kappa RF(R)g_{\mu\nu} \ , \tag{5.2.2}$$

where $F(R) \equiv df(R)/dR$, $\Delta_{\mu\nu} \equiv \nabla_\mu\nabla_\nu - g_{\mu\nu}\Box$ and the Einstein tensor is defined by $G_{\mu\nu} \equiv R_{\mu\nu} - (1/2)g_{\mu\nu}R$. GR is recovered by setting $f(R) = R$. In this model, the conservation of the energy-momentum tensor holds, just like in GR (it will become evident, in the development of the next section, that this is only at risk with a matter-geometry NMC).

For reference, note that the trace of equation (5.2.2),

$$2\kappa F(R)R = T - 6\kappa\Box F(R) + 4\kappa f(R) \ , \tag{5.2.3}$$

admits a de Sitter point, corresponding to a vacuum solution ($T = 0$) where the expansion is driven by a cosmological constant ($R = \text{const.}$). Then, a wider range of solutions than GR are admitted, as $R \neq 0$ when $T = 0$. In the former case, equation (5.2.3) reads

$$F(R)R - 2f(R) = 0 \ . \tag{5.2.4}$$

The function $f(R) = \alpha R^2$ satisfies this condition. Then, a de Sitter Universe can occur in Starobinsky model $f(R) = R + \alpha R^2$, in which the inflationary expansion lasts until αR^2 becomes smaller than the linear term R. In turn, one can also use the de Sitter point given by equation (5.2.4) in models for dark energy. A complete review of these theories is reported in Ref. [36].

The results used in the variational procedure are discussed in the following section, where we explore the $f(R)$ model with a non-trivial geometry-matter coupling included. Equation (5.2.2) is obtained from those results considering $f_2 = 1$; here, we just present some remarks to highlight — by comparison — the extended scenario that NMC theories provide.

5.2.2 Equivalence with a Scalar Field Theory

An interesting feature of $f(R)$ models is that they are equivalent to Jordan-Brans Dicke (JBD) theories and, performing a frame transformation, to scalar field theories [36]. To see this, we rewrite the $f(R)$ action as a function of an arbitrary field χ:

$$S = \int d^4x\sqrt{-g}\left[\kappa(f(\chi) + f'(\chi)(R - \chi)) + \mathcal{L}\right] \ . \tag{5.2.5}$$

Variation with respect to χ gives

$$f''(\chi)(R - \chi) = 0 \ , \tag{5.2.6}$$

which implies $\chi = R$ for a non-vanishing $f''(\chi)$. Inserting this result in equation (5.2.5), we obtain the original action (5.2.1).

Rewriting the action above in terms of the new field $\psi \equiv f'(\chi)$

$$S = \int d^4x\sqrt{-g}\left[\kappa\psi R - V(\psi) + \mathcal{L}\right] \ , \tag{5.2.7}$$

where the field potential is given by

$$V(\psi) = \kappa\left[\chi(\psi)\psi - f(\chi(\psi))\right] \ , \tag{5.2.8}$$

one obtains the standard form of the JBD action, without a kinetic term for the ψ-field. Action (5.2.7) obviously differ from the Einstein action by the introduction of an additional scalar field which is coupled to the Ricci scalar. In the JBD theory, the gravitational coupling is no longer a constant, but instead depends on the scalar field, which can vary from point to point in spacetime.

We can now write the $f(R)$ action as a scalar field theory in the Einstein frame, where the scalar curvature appears uncoupled. This can be done via a conformal transformation,

$$g_{\mu\nu} \to \tilde{g}_{\mu\nu} = \Omega^2 g_{\mu\nu} \ , \tag{5.2.9}$$

which implies the following relations [39]:

$$\sqrt{-g} = \Omega^{-4}\sqrt{-\tilde{g}_{\mu\nu}} \ , \tag{5.2.10}$$

$$R = \Omega^2\left[\tilde{R} - 6\Omega\tilde{\Box}\left(\frac{1}{\Omega}\right)\right] \ , \tag{5.2.11}$$

where $\tilde{\Box}$ denotes the D' Alembertian operator, defined from the metric $\tilde{g}_{\mu\nu}$ in the Einstein frame.

Substituting these results into the action (5.2.7), one obtains:

$$S = \int d^4x\sqrt{-\tilde{g}}\left[\kappa\frac{\psi}{\Omega^2}\tilde{R} - 6\kappa\frac{\psi}{\Omega}\tilde{\Box}\left(\frac{1}{\Omega}\right) - \frac{V(\psi)}{\Omega^4} + \frac{\mathcal{L}}{\Omega^4}\right] \ , \tag{5.2.12}$$

so that the conformal factor is defined through the condition

$$\Omega^2 = \psi \ . \tag{5.2.13}$$

Instead of maintaining the second term, we should derive an action endowed with a canonical kinetic term. Therefore, one integrates the covariant derivative and uses the metric compatibility relation, to get:

$$\begin{aligned} &\int \sqrt{\psi}\tilde{\Box}\left(\frac{1}{\sqrt{\psi}}\right)\sqrt{-\tilde{g}}d^4x \equiv \int \sqrt{\psi}\tilde{\nabla}_\mu\tilde{\nabla}^\mu\left(\frac{1}{\sqrt{\psi}}\right)\sqrt{-\tilde{g}}d^4x \\ &= \int \tilde{\nabla}_\mu\left[\sqrt{\psi}\tilde{\nabla}^\mu\left(\frac{1}{\sqrt{\psi}}\right)\sqrt{-\tilde{g}}\right]d^4x - \int \tilde{\nabla}_\mu\sqrt{\psi}\tilde{\nabla}^\mu\left(\frac{1}{\sqrt{\psi}}\right)\sqrt{-\tilde{g}}d^4x \ . \end{aligned} \tag{5.2.14}$$

The first integral may be dropped due to the divergence theorem, yielding

$$-6\kappa\int \sqrt{\psi}\tilde{\Box}\left(\frac{1}{\sqrt{\psi}}\right)\sqrt{-\tilde{g}}d^4x = -\frac{3\kappa}{2}\int \tilde{g}^{\mu\nu}\frac{\psi_{,\mu}\psi_{,\nu}}{\psi^2}\sqrt{-\tilde{g}}d^4x \ . \tag{5.2.15}$$

This can be cast in a simpler form, rescaling the field as

$$\phi \equiv \sqrt{3\kappa}\log(\psi) \ ; \tag{5.2.16}$$

then, the kinetic term directly reads:

$$\frac{3\kappa}{2}\frac{\psi_{,\mu}\psi_{,\nu}}{\psi^2} = \frac{3\kappa}{2}(\log\psi)_{,\mu}(\log\psi)_{,\nu} \equiv \frac{1}{2}\phi_{,\mu}\phi_{,\nu} \ . \tag{5.2.17}$$

We finally present the $f(R)$ action in the Einstein frame:

$$S = \int d^4x\sqrt{-\tilde{g}}\left[\kappa\tilde{R} - \frac{1}{2}\tilde{g}^{\mu\nu}\phi_{,\mu}\phi_{,\nu} - U(\phi) + e^{-2\frac{\phi}{\sqrt{3\kappa}}}\mathcal{L}\left(e^{-\frac{\phi}{\sqrt{3\kappa}}}\tilde{g}_{\mu\nu}, \Psi_M\right)\right] \ , \tag{5.2.18}$$

where $e^{-\frac{\phi}{\sqrt{3\kappa}}}\tilde{g}_{\mu\nu}$ is the physical metric, which is implicit in the interaction terms of general matter fields Ψ_M. Finally, U is the transformed scalar field potential, given by:

$$U = \Omega^{-4}V = \kappa\left[\frac{\chi(\psi)}{\psi} - \frac{f(\chi(\psi))}{\psi^2}\right] \ . \tag{5.2.19}$$

5.3 Non-minimally Coupled Theories

5.3.1 The Model

The generalized modified theory subject to our studies has an action of the form

$$S = \int [\kappa f_1(R) + f_2(R)\mathcal{L}]\sqrt{-g}d^4x \ , \tag{5.3.20}$$

where $f_i(R)$ (with $i = 1, 2$) are arbitrary functions of the scalar curvature R. In particular, a non-trivial $f_2(R)$ corresponds to a non-minimal coupling between matter and curvature [40] (see Refs. [41,42,43,44] for previous proposals and [45] for a review). The standard Einstein-Hilbert action is recovered by taking $f_2 = 1$ and $f_1 = R - 2\Lambda$.

Non-minimal coupled theories have the intriguing property that the four-divergence of the energy momentum tensor (of GR) is non-vanishing. Let us discuss this feature with more detail, starting by the derivation of the equation of motion in the metric formalism.

Varying the action (5.3.20) with respect to the metric yields

$$\delta S = \int \left[\frac{\kappa}{\sqrt{-g}} \frac{\delta(\sqrt{-g} f_1(R))}{\delta g^{\mu\nu}} + \frac{1}{\sqrt{-g}} \frac{\delta(\sqrt{-g} f_2(R)\mathcal{L})}{\delta g^{\mu\nu}} \right] \delta g^{\mu\nu} \sqrt{-g} d^4x \ , \tag{5.3.21}$$

with the energy-momentum tensor defined as

$$T_{\mu\nu} = -\frac{2}{\sqrt{-g}} \frac{\delta(\sqrt{-g}\mathcal{L})}{\delta g^{\mu\nu}} \ . \tag{5.3.22}$$

To simplify the expression above, we consider the following results:

$$\delta\sqrt{-g} = -\frac{1}{2}\sqrt{-g} g_{\mu\nu} \delta g^{\mu\nu} \ , \tag{5.3.23}$$

$$g^{\mu\nu} \delta R_{\mu\nu} = -\Delta_{\mu\nu} \delta g^{\mu\nu} \ , \tag{5.3.24}$$

$$\delta R = R_{\mu\nu} \delta g^{\mu\nu} + g^{\mu\nu} \delta R_{\mu\nu} \ , \tag{5.3.25}$$

where $R_{\mu\nu}$ is the Ricci tensor, $\Box = g^{\mu\nu}\nabla_\mu\nabla_\nu$ the d'Alembertian operator, and the Nabla operator is defined as before $\Delta_{\mu\nu} = \nabla_\mu\nabla_\nu - g_{\mu\nu}\Box$.

Then, the variation becomes

$$\begin{aligned} \delta S \quad &= \int \left[-\frac{1}{2}\kappa g_{\mu\nu} f_1(R) + \kappa f_1'(R) \frac{\delta R}{\delta g^{\mu\nu}} + f_2'(R)\mathcal{L} \frac{\delta R}{\delta g^{\mu\nu}} + f_2(R) \frac{1}{\sqrt{-g}} \frac{\delta(\sqrt{-g}\mathcal{L})}{\delta g^{\mu\nu}} \right] \delta g^{\mu\nu} \sqrt{-g} d^4x \\ &= \int \left[-\frac{1}{2}\kappa g_{\mu\nu} f_1(R) + \kappa f_1'(R) R_{\mu\nu} - \kappa f_1'(R)\Delta_{\mu\nu} + f_2'(R)\mathcal{L} R_{\mu\nu} - f_2'(R)\mathcal{L}\Delta_{\mu\nu} - \frac{1}{2} f_2(R) T_{\mu\nu} \right] \\ &\times \delta g^{\mu\nu} \sqrt{-g} d^4x \ . \end{aligned} \tag{5.3.26}$$

The corresponding field equations read

$$\left(F_1 + \frac{\mathcal{L}F_2}{\kappa} \right) R_{\mu\nu} = \frac{1}{2\kappa} f_2 T_{\mu\nu} + \Delta_{\mu\nu} \left(F_1 + \frac{\mathcal{L}F_2}{\kappa} \right) + \frac{1}{2} g_{\mu\nu} f_1 \ , \tag{5.3.27}$$

with $F_i \equiv df_i(R)/dR$.

One can write this in a more compact way:

$$F R_{\mu\nu} = \frac{1}{2} f_2 T_{\mu\nu} + \Delta_{\mu\nu} F + \frac{1}{2} g_{\mu\nu} \kappa f_1 \ , \tag{5.3.28}$$

defining $F = \kappa F_1(R) + F_2(R)\mathcal{L}(t)$. Or, making explicit the Einstein tensor,

$$F G_{\mu\nu} = \frac{1}{2} f_2 T_{\mu\nu} + \Delta_{\mu\nu} F + \frac{1}{2} g_{\mu\nu} \kappa f_1 - \frac{1}{2} g_{\mu\nu} R F \ . \tag{5.3.29}$$

Next, we calculate the covariant derivative of (5.3.29):

$$F' G_{\mu\nu} \nabla^\mu R = \frac{1}{2} F_2 T_{\mu\nu} \nabla^\mu R + \frac{1}{2} f_2 \nabla^\mu T_{\mu\nu} + \nabla^\mu \Delta_{\mu\nu} F + \frac{1}{2} g_{\mu\nu} \kappa F_1 \nabla^\mu R - \frac{1}{2} F g_{\mu\nu} \nabla^\mu R - \frac{1}{2} g_{\mu\nu} R F' \nabla^\mu R \ . \tag{5.3.30}$$

Writing explicitly $G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R$, we are left with

$$0 = \frac{1}{2} F_2 T_{\mu\nu} \nabla^\mu R + \frac{1}{2} f_2 \nabla^\mu T_{\mu\nu} - \frac{1}{2} g_{\mu\nu} \mathcal{L} F_2 \nabla^\mu R \ . \tag{5.3.31}$$

Hence, the final result:

$$\nabla_\mu T^{\mu\nu} = \frac{F_2}{f_2}(g^{\mu\nu}\mathcal{L} - T^{\mu\nu})\nabla_\mu R \ . \tag{5.3.32}$$

We have used Bianchi identity to make $\nabla^\mu G_{\mu\nu} = 0$; the metric compatibility, meaning $\nabla^\mu g_{\mu\nu} = 0$; and the fact that the covariant derivative of a scalar field is the normal derivative. We have also used the relation

$$\nabla^\mu \Delta_{\mu\nu} F_i(R) = (\Box\nabla_\nu - \nabla_\nu\Box)F_i(R) = R_{\mu\nu}\nabla^\mu F_i(R) \ , \tag{5.3.33}$$

to get equation (5.3.31). This arises directly from the definition of the Riemann curvature tensor, $\nabla_c\nabla_d X^a - \nabla_d\nabla_c X^a = R^a_{bcd}X^b$, contracting two indexes and noting that $\nabla_a\nabla_d F = \nabla_d\nabla_a F$ for a scalar function F, since $\nabla_j\nabla_i F = \nabla_j(\partial_i F) = \partial_i\partial_j F + \Gamma^m_{ji}\partial_m F$ and $\Gamma^m_{ij} = \Gamma^m_{ji}$ for the Levi-Civita connection.

Equation (5.3.32) leads to an extra force acting on a test particle, implying that it may deviate from its geodesic motion. To see this, we resort to the energy-momentum tensor (3.1.3) and consider the projection operator:

$$h_{\lambda\nu} = u_\lambda u_\nu - g_{\nu\lambda} \ , \tag{5.3.34}$$

from which one obtains $h_{\lambda\nu}u^\nu = 0$, as the fluid 4-velocity u_ν satisfies $u_\nu u^\nu = 1$ and $(\nabla_\mu u^\nu)u_\nu = 0$.

On one hand, note that

$$\nabla_\mu T^{\mu\nu} = (\rho + p)_{,\mu}u^\mu u^\nu + (\rho + p)\nabla_\mu(u^\mu u^\nu) - p_{,\mu}g^{\mu\nu} \ ; \tag{5.3.35}$$

by contracting this with the projection operator, one obtains:

$$h_{\lambda\nu}[\nabla_\mu T^{\mu\nu}] = (\rho + p)(\nabla_\mu u^\nu)u^\mu g_{\nu\lambda} - p_{,\mu}g^{\mu\nu}(u_\lambda u_\nu - g_{\nu\lambda}) \ . \tag{5.3.36}$$

On the other hand, contracting the r.h.s. of equation (5.3.32) with $h_{\lambda\nu}$, one gets:

$$h_{\lambda\nu}\left[\frac{F_2}{f_2}(g^{\mu\nu}\mathcal{L} - T^{\mu\nu})\nabla_\mu R\right] = \frac{F_2}{f_2}(\mathcal{L} + p)h_{\lambda\nu}\ .\nabla^\nu R \tag{5.3.37}$$

Gathering these results, one encounters the following expression:

$$(\rho + p)u^\mu(\nabla_\mu u_\lambda) - (\nabla^\nu p)h_{\lambda\nu} = \frac{F_2}{f_2}(\mathcal{L} + p)h_{\lambda\nu}\nabla^\nu R \ , \tag{5.3.38}$$

or, equivalently,

$$\frac{du_\lambda}{ds} - \Gamma^\alpha_{\mu\lambda}u_\alpha u^\mu = f_\lambda \ , \tag{5.3.39}$$

where

$$f_\lambda = \frac{1}{\rho + p}\left[\frac{F_2}{f_2}(\mathcal{L} + p)\nabla^\nu R + \nabla^\nu p\right]h_{\lambda\nu} \ . \tag{5.3.40}$$

Thus, additionally to the usual gradient pressure of the fluid, an extra contribution arises due to the NMC. This is zero, provided the matter Lagrangian is assumed to be $\mathcal{L} = -p$.

5.3.2 Equivalence with Scalar-Tensor Theories

We can rewrite NMC theories with an action with two scalar fields[2] [46]. To see this, one again establishes the equivalence between action (5.3.20) and a general scalar-tensor theory, as the Jordan-Brans-Dicke theory:

$$S = \int d^4x\sqrt{-g}\left[\psi R - \frac{\omega}{2\psi}g^{\mu\nu}\partial_\mu\psi\partial_\nu\psi - V(\chi,\psi) + f_2(\chi)\mathcal{L}\right] , \tag{5.3.41}$$

where ω is a parameter of the theory that can be chosen to fit observations.

Starting with the action:

$$S = \int d^4x\sqrt{-g}[\kappa f_1(\chi) + \psi(R-\chi) + f_2(\chi)\mathcal{L}] , \tag{5.3.42}$$

variation with respect to ψ and χ gives, respectively:

$$\chi = R ; \tag{5.3.43}$$

$$\psi = \kappa f_1'(\chi) + f_2'(\chi)\mathcal{L} . \tag{5.3.44}$$

This implies that χ and ψ are independent, provided $\mathcal{L}$ and $f_2'(\chi)$ are non-zero. Substituting into (5.3.42), one recovers the action of our model (5.3.20).

We can rewrite the action (5.3.42) in the form of a JBD type theory (5.3.41) with $\omega = 0$:

$$S = \int d^4x\sqrt{-g}\,[\psi R - V(\chi,\psi) + f_2(\chi)\mathcal{L}] , \tag{5.3.45}$$

for a potential given by

$$V(\chi,\psi) = \psi\chi - \kappa f_1(\chi) . \tag{5.3.46}$$

This corresponds to a JBD theory in the Jordan frame, where the curvature appears linearly coupled to a function of scalar fields. Again, one can perform a conformal transformation, so that the curvature appears decoupled, yielding the action in the so-called Einstein frame.

Thus, using the conformal transformation (5.2.9) and inserting the previous results (5.2.10-5.2.11) into the action (5.3.45), one obtains:

$$S = \int d^4x\sqrt{-\tilde{g}}\left[\frac{\psi}{\Omega^2}\tilde{R} - 6\frac{\psi}{\Omega}\tilde{\Box}\left(\frac{1}{\Omega}\right) - U + \frac{f_2(\chi)}{\Omega^4}\mathcal{L}\right] , \tag{5.3.47}$$

and, consequently, the Einstein frame is defined by taking $\Omega^2 = \psi$,

$$S = \int d^4x\sqrt{-\tilde{g}}\left[\tilde{R} - 6\sqrt{\psi}\tilde{\Box}\left(\frac{1}{\sqrt{\psi}}\right) - U + \frac{f_2(\chi)}{\psi^2}\mathcal{L}\right] , \tag{5.3.48}$$

where one defines

$$U(\chi,\psi) = \Omega^{-4}V(\chi,\psi) = \frac{\chi}{\psi} - \kappa\frac{f_1(\chi)}{\psi^2} . \tag{5.3.49}$$

[2]The NMC coupling introduces an extra degree of freedom, hence we add one more field, relative to what is done in the case of $f(R)$ theories.

One now attempts to recast the action in terms of two other scalar fields, endowed with a canonical kinetic term. For this purpose, one uses the result (5.2.14) to rewrite the second term of the action as

$$-6\int\sqrt{\psi}\tilde{\Box}\left(\frac{1}{\sqrt{\psi}}\right)\sqrt{-\tilde{g}}d^4x=-\frac{3}{2}\int\tilde{g}^{\mu\nu}\frac{\psi_{,\mu}\psi_{,\nu}}{\psi^2}\sqrt{-\tilde{g}}d^4x\ . \tag{5.3.50}$$

To put this in the form of a canonical kinetic term, one aims at writing something like

$$\frac{3}{2}\frac{\psi_{,\mu}\psi_{,\nu}}{\psi^2}=\frac{3}{2}(\log\psi)_{,\mu}(\log\psi)_{,\nu}\equiv 2\sigma_{ij}\varphi^i_{,\mu}\varphi^i_{,\nu}\ , \tag{5.3.51}$$

where we can interpret σ_{ij} $(i,j=1,2)$ as the metric of the two-dimensional space of scalar fields [46], and φ^1,φ^2 are the two new scalar fields. An easy identification that satisfies (5.3.51) is given by:

$$\varphi^1=\frac{\sqrt{3}}{2}\log\psi\ ,\quad \varphi^2=\chi\ , \tag{5.3.52}$$

and the field metric

$$\sigma_{ij}=\begin{pmatrix}1&0\\0&0\end{pmatrix}\ , \tag{5.3.53}$$

indicating that only the φ^1 field has a kinetic term. Despite this, $\varphi^2=\chi$ is a distinct degree of freedom, because one cannot write the potential $U(\chi,\psi)$ in terms of one scalar field alone:

$$U(\varphi^1,\varphi^2)=\exp\left(-\frac{2\sqrt{3}}{3}\varphi^1\right)\left[\varphi^2-\kappa f_1(\varphi^2)\exp\left(-\frac{2\sqrt{3}}{3}\varphi^1\right)\right]\ . \tag{5.3.54}$$

In the trivial case where $f_2(R)=1$ or $\mathcal{L}=0$, one gets $\psi=\kappa f_1'(\chi)$, implying $\varphi^1\propto\log f_1'(\varphi^2)$: one degree of freedom is lost and the potential may be written as a function of just one of the fields.

Introducing the redefinition of the two scalar fields, the action reads:

$$S=\int d^4x\sqrt{-\tilde{g}}\left[\tilde{R}-2\tilde{g}^{\mu\nu}\sigma_{ij}\varphi^i_{,\mu}\varphi^j_{,\nu}-U(\varphi^1,\varphi^2)+f_2(\varphi^2)e^{-\frac{4}{\sqrt{3}}\varphi^1}\mathcal{L}\left(e^{-\frac{2}{\sqrt{3}}\varphi^1}\tilde{g}_{\mu\nu},\Psi_M\right)\right]\ , \tag{5.3.55}$$

where Ψ_M denotes all matter fields not related to the scalar curvature and the matter Lagrangian density, as φ^1 and φ^2 are.

Discussion

We should notice that the equivalence we have just established is not to an ordinary JBD theory, due to the presence of the scalar-matter coupling in the action (5.3.42). Although the ordinary JBD theory in the Einstein frame has a non-minimal scalar-matter coupling similar with the term $f_2(\chi)\mathcal{L}$ (setting $f_2=1$ in equation 5.3.55, we are still left with an exponential function of the scalar field multiplying the matter Lagrangian), there is an important difference: choosing the Einstein frame, the units of time, length and mass are not constant, but scale with the appropriate powers of the conformal factor defining this frame. With a non-trivial $f_2(R)$-function, however, this cannot solve the problem, because the NMC is already explicit in the Jordan frame — see equation (5.3.42) —, where there is no such scaling.

Although $f(R)$ theories with a non-minimal curvature-matter coupling cannot be cast into the standard scalar-tensor theories, we were able to pursue an equivalence with two scalar fields. This equivalence was established in the Einstein frame, through a conformal transformation, where the curvature and the metric become minimally coupled, i.e., the gravitational sector of the action reduces to Einstein's canonical form.

Going from one frame to another, we end up fixing a new canonical set of scalar fields and a new potential. It is widely discussed in the literature the equivalence of these two frames, which is not obvious because we do not know if there exists or not a conformal invariance of the underlying modified theory [39,47].

The argument of the physical equivalence between the Jordan and Einstein frames is that the conformal transformation of the metric is not just a mathematical artifact, but instead scales the units of time, length and mass with appropriate powers of the conformal factor, Ω. Then, since physics must be invariant under a change of units, it should be invariant under conformal transformations *with the associated rescaling of units*, leading to the same physical predictions in both frames. This interpretation is as follows: the metric scales as $\tilde{g}_{\mu\nu} = \Omega^2 g_{\mu\nu}$; hence, the times and lengths scale with Ω, $\tilde{dt} = \Omega dt$ and $\tilde{dx}^i = \Omega dx^i$. Then, from dimensional analysis, the masses scale as $\tilde{m} = \Omega^{-1} m$.

In the Jordan frame, the gravity effective coupling varies, but the masses of the elementary particles do not. In the Einstein frame, this is interchanged. Consider, for example, the proton mass, m_p. On the Jordan frame, it has a constant value; while, in the Einstein frame, it scales as $\tilde{m}_p = \Omega^{-1} m_p$. However, what we measure in an experiment is the *ratio* between the proton mass and an arbitrarily chosen mass unit, which would yield the same value in both frames.

This is just the tip of the iceberg in what concerns the discussion about the designated *equivalence*, because — in the Einstein frame — not only the masses of elementary particles and the mass units, but also the coupling constants of nongravitational physics vary with the scalar fields [48]. The choice of the conformal frame has been widely studied in the context of Cosmology, as well [47].

5.3.3 The choice of the Lagrangian

With the introduction of the NMC, the Lagrangian density of matter appears explicitly in the field equations (5.3.28). In the standard cosmological scenario, where matter behaves as a perfect fluid, the choice of this quantity is not unique. It is therefore fundamental to understand if the classical equivalence between different Lagrangian densities of a perfect fluid holds, in the presence of a non-trivial matter-geometry coupling.

Perfect fluids, which we introduced in section 3.1, are described locally by various thermodynamical variables:

$$n = \text{particle number density} \ , \tag{5.3.56}$$

$$\rho = \text{energy density} \ , \tag{5.3.57}$$

$$p = \text{pressure} \ , \tag{5.3.58}$$

$$T = \text{temperature} \ , \tag{5.3.59}$$

$$s = \text{entropy per particle} \ . \tag{5.3.60}$$

By definition, the energy-momentum tensor for a perfect fluid has the form

$$T^{\mu\nu} = (p+\rho)u^\mu u^\nu - p g^{\mu\nu} \ , \tag{5.3.61}$$

where u^μ is the unit 4-velocity of the fluid.

In GR, the perfect fluid action functional [49] incorporates the energy tensor (5.3.61); the required equations of motion, namely $(nu^\mu)_{;\mu} = 0$, which expresses conservation of particle number; and the First Law of Thermodynamics

$$d\rho = \mu dn + nT ds \ , \tag{5.3.62}$$

for a perfect fluid with equation of state $\rho(n,s)$.

This action is introduced as a function of a particle number flux vector

$$J^\mu = \sqrt{-g} n u^\mu \ , \tag{5.3.63}$$

that is, the fluid 4-velocity is defined by

$$u^\mu = J^\mu / |J| \ , \tag{5.3.64}$$

where

$$|J| = \sqrt{g_{\mu\nu} J^\mu J^\nu} \tag{5.3.65}$$

is the magnitude of J^μ and the particle number density is given by

$$n = |J| / \sqrt{-g} \ . \tag{5.3.66}$$

The perfect fluid action is then presented in terms of Lagrangian coordinates α^A and spacetime scalars ψ, θ and β_A $(A = 1,2,3)$, which allow us to enforce constraints on the physical variables (they act as Lagrange multipliers):

$$S_m = \int d^4x \left[\sqrt{-g}\rho(n,s) + J^\mu(\psi_{,\mu} + s\theta_{,\mu} + \beta_A \alpha^A_{,\mu}) \right] \ . \tag{5.3.67}$$

Recalling the definition of the energy-momentum tensor,

$$T^{\mu\nu} = \frac{2}{\sqrt{-g}} \frac{\delta(\sqrt{-g}\mathcal{L}_m)}{\delta g_{\mu\nu}} \ , \tag{5.3.68}$$

the variation with respect to the metric yields

$$\begin{aligned}
T^{\mu\nu} \quad &= \frac{2}{\sqrt{-g}}\left[\rho\frac{\delta(\sqrt{-g})}{\delta g_{\mu\nu}} + \sqrt{-g}\frac{\partial\rho}{\partial n}\left(\frac{\partial n}{\partial|J|}\frac{\partial|J|}{\partial g_{\mu\nu}} + \frac{\partial n}{\partial\sqrt{-g}}\frac{\delta(\sqrt{-g})}{\delta g_{\mu\nu}}\right)\right] \\
&= \frac{2}{\sqrt{-g}}\left[\frac{\rho}{2}\sqrt{-g}g^{\mu\nu} + \sqrt{-g}\frac{\partial\rho}{\partial n}\left(\frac{1}{\sqrt{-g}}\frac{J^\mu J^\nu}{2|J|} - \frac{|J|}{(\sqrt{-g})^2}\frac{\sqrt{-g}}{2}g^{\mu\nu}\right)\right] \\
&= \rho g^{\mu\nu} + n\frac{\partial\rho}{\partial n}(u^\mu u^\nu - g^{\mu\nu}) \\
&= \rho u_\mu u_\nu + \left(n\frac{\partial\rho}{\partial n} - \rho\right)(u_\mu u_\nu - g_{\mu\nu}) \ ,
\end{aligned} \tag{5.3.69}$$

which, comparing with (5.3.61), defines the pressure as

$$p = n\frac{\partial\rho}{\partial n} - \rho \ . \tag{5.3.70}$$

This definition is in agreement with the First Law of Thermodynamics, written above, which shows that an equation of state for a perfect fluid can be specified by giving the energy density $\rho(n,s)$ as a function of the number density and entropy per particle. We directly identify $\mu = \partial\rho/\partial n$ with the usual definition for the chemical potential, $\mu = (\rho+p)/n$.

The variation with respect to $J^\mu, \psi, \theta, s, \alpha^A$ and β_A provides the following equations of motion:

$$\frac{\delta S}{\delta J^\mu} = \mu u_\mu + \psi_{,\mu} + s\theta_{,\mu} + \beta_A\alpha^A_{,\mu} = 0 \ , \tag{5.3.71}$$

$$\frac{\delta S}{\delta\psi} = -J^\mu_{,\mu} = 0 \ , \tag{5.3.72}$$

$$\frac{\delta S}{\delta\theta} = -(sJ^\mu)_{,\mu} = 0 \ , \tag{5.3.73}$$

$$\frac{\delta S}{\delta s} = \sqrt{-g}\frac{\partial\rho}{\partial s} + \theta_{,\mu}J^\mu = 0 \ , \tag{5.3.74}$$

$$\frac{\delta S}{\delta\alpha^A} = -(\beta_A J^\mu)_{,\mu} = 0 \ , \tag{5.3.75}$$

$$\frac{\delta S}{\delta\beta_A} = \alpha^A_{,\mu}J^\mu = 0 \ . \tag{5.3.76}$$

The first relationship, equation (5.3.71), provides the velocity-representation of the 4-velocity. The second equation (5.3.72) reflects particle conservation, while the third (5.3.73) translates the entropy exchange constraint. Equation (5.3.74) provides the identification of the temperature

$$T = \frac{1}{n}\frac{\partial\rho}{\partial s}\bigg|_n = -\theta_{,\mu}u^\mu \ , \tag{5.3.77}$$

after replacing the definition of the density flux and comparing with the First Law of Thermodynamics.

Equation (5.3.75) reflects the constancy of β_A along the fluid lines; conversely, equation (5.3.76) restricts the fluid 4-velocity to be directed along the flow lines of constant α^A.

Now taking into account equation (5.3.71) and the definitions above, the action (5.3.67) reduces to the on-shell Lagrangian density $\mathcal{L}_{m(1)} = -p$ with the action given by:

$$S_m = -\int d^4x\sqrt{-g}p \ , \tag{5.3.78}$$

which yields the same equations of motion (5.3.71-5.3.76) if one considers the pressure is functionally dependent on the previous spacetime fields $\phi, s, \theta, \beta_A, \alpha_A$ and on the current density J^μ.

The on-shell degeneracy of the Lagrangian densities arises from adding up surface integrals to the action. Suppose we start with:

$$S = \int d^4x[\sqrt{-g}\rho(n,s) - \psi J^\mu_{,\mu} - \theta(sJ^\mu)_{,\mu} - \alpha^A(\beta_A J^\mu)_{,\mu}] \; ; \tag{5.3.79}$$

noticing that this action reproduces the equations of motion (5.3.71-5.3.76), it reduces to

$$S = \int d^4x\sqrt{-g}\rho(n,s) \; , \tag{5.3.80}$$

which, again, reproduces the energy-momentum tensor of the fluid (with ρ independent of the spacetime fields).

Hence, we conclude that the Lagrangian choice is not unique. Adding up to $\mathcal{L}_{m(2)} = \rho$, it can be shown that the on-shell Lagrangian $\mathcal{L}_{m(3)} = na$, where a is the physical free energy $a(n,T)$, can reproduce the correct stress-energy tensor of a perfect fluid.

Having this in mind, one now describes how this treatment could be generalized for a non-trivial NMC [50]. This should only affect terms in equation (5.3.67) which are minimally coupled already.

Defining (for simplicity) the general set of thermodynamic potentials and Lagrangian multipliers as $\phi_\mu = \psi_{,\mu} + s\theta_{,\mu} + \beta_A\alpha^A_{,\mu}$, the matter action becomes:

$$S'_m = \int d^4x(\sqrt{-g}f_2(R)\rho + J^\mu\phi_\mu) \; , \tag{5.3.81}$$

where the current density term remains uncoupled (except for the use of the metric to contract indexes). The equations of motion (5.3.72), (5.3.73), (5.3.75) and (5.3.76) are unchanged, while (5.3.71) and (5.3.74) read:

$$\frac{\delta S'}{\delta J^\mu} = f_2\mu u_\mu + \phi_\mu = 0 \; , \tag{5.3.82}$$

and

$$\frac{\delta S'}{\delta s} = \sqrt{-g}f_2\frac{\partial\rho}{\partial s} + \theta_{,\mu}J^\mu = 0 \; . \tag{5.3.83}$$

Thus, the non-minimal coupling of curvature with matter is reflected in both the velocity and the temperature identification, as now:

$$T = \frac{1}{n}\frac{\partial\rho}{\partial s}\bigg|_n = -\frac{1}{f_2}\theta_{,\mu}u^\mu \; . \tag{5.3.84}$$

By substituting the modified equations of motion into action (5.3.81), the on-shell Lagrangian density $\mathcal{L}_{m(1)} = -p$ can be read, as in GR. By including extra surface integrals, a similar procedure also yields $\mathcal{L}_{m(2)} = \rho$ and $\mathcal{L}_{m(3)} = na$.

Even though the action can still adopt distinct on-shell Lagrangian densities — which generate the same energy-momentum tensor and the same equations of motion —, they are no longer equivalent in the presence of the NMC. This is because, not only the equations of motion for fields describing the perfect fluid, but also the gravitational field equations must remain invariant with this choice.

Different predictions for nongeodesic motion result from different forms of the gravitational equations. It is clear from expression (5.3.40) that this depends on the form of the Lagrangian, hence the equivalence between the on-shell densities $\mathcal{L}_{m(i)}$ and the original one $\mathcal{L}_m$ is broken and one cannot choose arbitrarily between the available forms.

Nevertheless, this bare quantity defined in equation (5.3.67) should not appear in the non-conservation law (5.3.32): when deriving the energy-momentum tensor, we vary the action with respect to the fields which are coupled to the metric, which is not the case of the current term. We are thus left with $\mathcal{L}_{m(2)} = \rho$.

In the case of matter which becomes non-relativistic as the temperature drops, behaving as pressureless dust $\rho \gg p$, it appears unnatural to take a vanishing quantity $\mathcal{L}_{m(1)} = -p$ as the Lagrangian density. Moreover, in our previous discussion of the functional description of a perfect fluid, we found out pressure is not an independent quantity. A more rigorous formulation to characterize a dust distribution is then that of an isentropic fluid with an equation of state of the form $\rho(n) = n\mu$, with a constant chemical potential[3].

This concludes the discussion that justifies the choice of $\mathcal{L} = \rho$ instead of $\mathcal{L} = -p$ for the matter Lagrangian density, in the presence of a non-minimal geometry-matter coupling.

5.3.4 Thermodynamic Interpretation

As derived in section 5.3.1, the presence of the NMC gives rise to the non-conservation of the effective energy-momentum tensor. This can be interpreted as an energy exchange between matter and gravity, due to the established equivalence between these theories and two coupled scalar fields. Therefore, we could simply proceed by extending equation (3.1.9) to non-adiabatic transformations, regarding the non-zero contribution in the modified covariant law for $T^{\mu\nu}$ as an entropy flow between those fields. We could rewrite equation (5.3.29) as follows:

$$G_{\mu\nu} = \hat{k}\left(\widehat{T}_{\mu\nu} + T_{\mu\nu}\right) \ , \tag{5.3.85}$$

where the hat-components correspond to modifications introduced by the non-vanishing $f_i(R)$ functions,

$$\widehat{T}_{\mu\nu} = \frac{2}{f_2}\Delta_{\mu\nu}(\kappa F_1 + \mathcal{L}F_2) + \frac{1}{f_2}g_{\mu\nu}\left[\kappa f_1 - R(\kappa F_1 + \mathcal{L}F_2)\right] \tag{5.3.86}$$

and

$$\hat{k} = \frac{f_2}{2(\kappa F_1 + \mathcal{L}F_2)} \ . \tag{5.3.87}$$

By using the Bianchi identities on equation (5.3.85), one concludes that, although $T_{\mu\nu}$ is no longer conserved, the modified theory with a non-trivial $f_2(R)$ has a *new* stress-energy tensor which is conserved[4].

[3] We obtain this result from equation (5.3.70), considering a negligible pressure.

[4] The generalized energy-momentum tensor of the modified theory is implicitly identified with the r.h.s. of (5.3.85), which is covariantly conserved. The explicit identification (if possible) requires rewriting this term as $kT'_{\mu\nu}$, where k is interpreted

This hints toward a reinterpretation of the effective non-conservation of $T_{\mu\nu}$ as an energy flow between the gravitational and the matter fields. These considerations are only valid if we extend the Standard Cosmology, whose energy equations are adiabatic and reversible. Consider the following analogy: a student examines the dynamics of a snooker table; then, at some point, someone tilts the system without him realizing. The equations of motion for the system in the plane would thus fail. He would only see the balls fall off after they climb. To incorporate this into the 2D equations, he would evoke a somehow magical current that exchanges energy with the snooker balls. This standard thermodynamic view is plausible (if we cannot look from a higher dimensional viewpoint), but lacks meaning.

The extension of thermodynamics to general relativity is not standard and the generalization of the fundamental laws must be used with care, but there is no apparent reason why these should not be applied to cosmic expansion. We should be able to formulate an effective energy conservation equation — the equivalent to the first law — even in a modified gravity theory with the property (3.1.9). This is possible, identifying a non-vanishing variation of the entropy — which, if not decreasing, is compatible with the second law of thermodynamics.

Recently, the non-conservation of the energy-momentum tensor in the presence of the NMC has been interpreted in the framework of open systems. In the cosmological context, we can give a meaning to this additional contribution: one can consider that it gives rise to an irreversible energy flow from the gravitational field to matter constituents, which allows for both particle and entropy production in the early Universe. This is interesting to our work, because if this extra entropy can create more particles than the Standard Cosmology allows, the argument that leads to the constancy of the asymmetry parameter η_S — which relies on an isentropic cosmological expansion — may fail.

Even before $f(R)$ theories with a NMC were a current topic of investigation, Prigogine [51] showed that the generalization of the concept of adiabatic transformations from closed to open systems — where irreversible thermodynamic processes are included — leads to a reinterpretation of the energy-momentum tensor which takes into account matter creation as a source of internal energy. This model inherently relies on an enlargement of the traditional cosmology, since the GR expression for energy conservation does not provide a framework for particle production (Einstein's equations are solved assuming an equation of state depending on two physical variables only: the energy-density ρ and the pressure p). As stated by Prigogine, a model with an alternative cosmology could quite naturally serve this phenomenological approach: recalling the energy conservation equation in GR, $\dot{\rho} + 3H(\rho + p) = 0$, it suffices to have a non-vanishing r.h.s. to discuss the existence of some supplementary effective pressure *which should be related to particle creation*. The latter needs, however, additional justification. If we go along this line of reasoning, we will need to clarify the proper use of thermodynamics, mainly the modifications to the first and second laws, which are necessary in the context of open thermodynamic systems.

as an effective coupling constant, satisfying $\nabla^\mu k = 0$. This directly implies $\nabla^\mu T'_{\mu\nu} = 0$, establishing $T'_{\mu\nu}$ as the covariantly conserved energy-momentum tensor of the modified theory.

More recently, Harko [52] studied how the generalized (non-)conservation equations in gravity theories with a non-minimal geometry-matter coupling can be interpreted as the source of these irreversible matter creation processes[5]. This particular interpretation constitutes an attractive procedure to discuss in this work, since the presence of a NMC destroys the usual conservation laws, hence the thermodynamics of the *cosmological fluid* need to be reformulated accordingly. Putting in perspective, it poses several advantages:

- It can shed light on the problem of the initial conditions, giving an explanation for the origin of cosmological entropy and providing the entropy burst accompanying the production of matter;
- In spite of the non-conservation of the energy-momentum tensor of GR (from which we can extract the physical components characterizing the thermodynamic behaviour of the Universe), we are able to formulate an effective conservation equation, that includes not only energy in the form of normal pressure, like the one carried by the photons, but also an entropy variation, for which the gravitational field acts as a source.
- Finally, this framework allows us to compute the thermodynamic quantities characterizing the model for irreversible matter production, like the creation pressure, the particle creation rate and the entropy production rate [52], as these are completely determined by the gravitational action.

In this framework, one thinks of the Universe as a two-fluid system in which the cosmological entropy is closely related to the variation of the number of particles. These are added to spacetime, through an energy flow from the gravitational to matter fields, which exists due to the presence of the NMC. Accordingly, the object of this study is the open matter-gravitation system.

In what concerns baryogenesis, this means that we can change the number of baryons while the entropy rate is large enough to create particles. But we do not see any experimental evidence of decays that change the baryon number; so, below some scale, there should not be any large variation of the cosmological entropy. Therefore, computing the timescale variation of this *gravitational entropy* (the entropy necessary to produce matter from the gravitational field) will be the ultimate test of any model that attempts to generate the right amount of baryon asymmetry while admitting alternative cosmologies that alter the usual adiabatic conservation law of the energy-momentum tensor.

Thermodynamics of Matter Creation

Let us consider a volume V containing N particles. For a closed system, N is constant. The corresponding thermodynamic law expressing the conservation of the internal energy E is given by:

$$dE = dQ - pdV \ , \tag{5.3.88}$$

[5] After all, the r.h.s. of equation (5.3.32) is non-vanishing, possibly leading to the analysis we have just discussed.

where dQ is the heat received by the system during time dt.

We already studied the conservation law for the energy-momentum tensor in GR. We have concluded that Einstein equations are adiabatic and reversible:

$$d(\rho V) + pdV = 0 \ . \tag{5.3.89}$$

Now we want to extend the concept of adiabatic transformation from closed to open systems. Then, we need to include in (5.3.88) a term that explicitly takes into account the variation of the number of particles:

$$d(\rho V) = dQ - pdV + (h/n)d(nV), \quad \text{where} \quad h = \rho + p \tag{5.3.90}$$

is the enthalpy (per unit volume), ρ is the energy density and n is the particle number density.

We restrict our treatment to adiabatic transformations, $dQ = 0$, thus neglecting proper heat transfer processes in the cosmological system. Note that we usually write the First Law with μdN to account for the variation of the number of particles, in the presence of a chemical force μ. By writing this in terms of the enthalpy, we are saying that the number of particles can vary due to entropy production, as well:

$$h = \mu n + Ts \ ; \tag{5.3.91}$$

this is only true for irreversible processes, where entropy is allowed to vary even for adiabatic transformations.

The generalization of the First Law to open systems thus reads:

$$\frac{d}{dt}(\rho a^3) + p\frac{d}{dt}a^3 = \frac{h}{n}\frac{d}{dt}(na^3) \ , \tag{5.3.92}$$

for a system containing N particles in a comoving volume $V = a^3$. In such a transformation, *the "heat" received by the system is entirely due to the change in the number of particles* [51].

Equation (5.3.92) can be written as

$$\dot{\rho} + 3H(\rho + p) = \frac{\rho + p}{n}(\dot{n} + 3Hn) \ . \tag{5.3.93}$$

Next, we write the 00−component of equation (5.3.32). This reads:

$$\dot{\rho} + 3H(\rho + p) = \frac{d[\ln f_2(R)]}{dt}(\mathcal{L} - \rho) \ . \tag{5.3.94}$$

By comparing these two equations, we can interpret the energy conservation in $f(R)$ theories with a NMC as describing matter creation in a FRW Universe:

$$\dot{n} + 3Hn = \frac{n}{\rho + p}\frac{d[\ln f_2(R)]}{dt}(\mathcal{L} - \rho) \ . \tag{5.3.95}$$

This equation gives the particle number variation with time in an expanding Universe, with the r.h.s. being identified with Γn, where Γ is the particle creation rate. This clarifies the mechanism of matter creation, although details are beyond the aim of our work.

Gravitational Entropy Production

We turn now to the Second Law of Thermodynamics. This induces us to decompose the entropy change into an entropy flow d_eS and an entropy creation d_iS [53]:

$$dS = d_eS + d_iS \ , \quad \text{with} \quad d_iS \geq 0 \ . \tag{5.3.96}$$

This decomposition follows from the result that, comparing the evolution of a system between two thermodynamic states by a reversible path or by an irreversible one, we obtain $\int_{\text{irrev}} dQ/T < \int_{\text{rev}} dQ/T$. This means that, for an irreversible process, $\int_{\text{irrev}} dQ/T$ does not contain all contributions to the entropy change (due to the disorder created by spontaneity). This allows us to write

$$dS = \frac{dQ}{T} + dS_i \ , \tag{5.3.97}$$

which we identify with (5.3.96). Here dS_i denotes the entropy production due to spontaneous activity.

For a reversible process, $d_iS = 0$ and $dS = dQ/T$, so the entropy change is entirely due to a flow of heat in or out of the system. For an isolated system, where there is no external flow of heat, we have $dQ = 0$; therefore

$$dS = dS_i \geq 0 \ , \tag{5.3.98}$$

where the equality holds for a reversible process and the inequality holds for a spontaneous, irreversible one (which, in our case, is the only way of entropy production).

From the total differential of the entropy, $TdS = dE + pdV - \mu dN$, and gathering the previous results, we get

$$Td_iS = TdS = \frac{h}{n}d(nV) - \mu d(nV) = T\frac{s}{n}d(nV) \geq 0 \ , \tag{5.3.99}$$

to account for the entropy production due to matter creation, using (5.3.91). From this equation, we obtain the time variation of the entropy

$$\frac{dS}{dt} = \frac{S}{n}(\dot{n} + 3Hn) \geq 0 \ , \tag{5.3.100}$$

in an expanding Universe.

Finally, making use of equation (5.3.95), giving the particle number balance in this class of gravity theories, we end up with:

$$\frac{1}{S}\frac{dS}{dt} = \frac{1}{\rho + p}\frac{d[\ln f_2(R)]}{dt}(\mathcal{L} - \rho) \geq 0 \ , \tag{5.3.101}$$

which enables us to compute the timescale variation of the entropy production, for a designated form of $f_2(R)$.

Chapter 6

Gravitational Baryogenesis

Finally, we attempt to construct a model for Baryogenesis resorting to modifed gravity theories. For this purpose, we aim at reinterpreting Cohen's effective coupling using a scalar quantity that characterizes the gravitational interaction. Once we identify this with the Ricci scalar, the baryon asymmetry can be generated quite naturally in the course of evolution of our Universe.

6.1 Motivation

As we have been discussing, the baryon asymmetry in the Universe is inferred from the baryon-to-photon ratio. This allows us to determine the baryon-to-entropy ratio, which is preserved since baryon violating forces drop out of equilibrium. Quantitatively (3.3.64), it corresponds to:

$$\eta_S^{\mathrm{obs}} \equiv \frac{n_B}{s} \lesssim 9 \times 10^{-11} \ . \tag{6.1.1}$$

The fact that $n_b \gg n_{\bar{b}}$ means that η_S is also the ratio of the net baryon number, n_B, to the entropy density, s. What remains a mystery is how the baryon asymmetry was generated.

Avoiding the out-of-equilibrium scenario, Cohen and Kaplan showed that one can generate η while preserving thermal equilibrium, in an expanding Universe that admits a dynamical violation of CPT. This was accomplished by the introduction of an effective coupling between the baryonic current and a scalar field, $(\partial_\mu \phi) J^\mu$.

Following their work, Davoudiasl [54] considered an identical coupling, but instead of some scalar field, he used the Ricci scalar curvature R:

$$\frac{1}{M_*^2} \int d^4x \sqrt{-g} (\partial_\mu R) J^\mu \ , \tag{6.1.2}$$

where M_* is the cutoff scale of the effective theory. It is natural to expect such an operator in the low-energy effective field theory of quantum gravity, if not forbidden by some symmetry of the theory.

Notice that this interaction eventually becomes vanishingly small for a matter or radiation dominated Universe, since the scale factor behaves as a power-law, $a(t) \sim t^{2/3(1+\omega)}$, thus implying that the scalar curvature drops as t^{-2} and $\dot{R} \sim t^{-3}$. This is of fundamental importance, as interaction (6.1.2) can be effective only for a temporary period in the primordial stages of evolution; otherwise, the Universe would not evolve into a CPT-invariant ground state.

To generate a baryon asymmetry via interaction (6.1.2), we still need to require the existence of B-violating processes. We denote the temperature at which B-violating forces decouple from the thermal bath by T_D. Then, the B-asymmetry is generated as follows:

1. In an expanding Universe, the interaction in (6.1.2) — provided $\dot{R}$ is non-vanishing — gives opposite sign energy contributions that differ for particles and antiparticles, and thereby dynamically violates CPT;
2. This modifies thermal equilibrium distributions in a similar way as a chemical potential $\mu_b \sim \dot{R}/M_*^2 = -\mu_{\bar{b}}$ would, driving the Universe towards a nonzero equilibrium B-asymmetry, as permitted by the B-violating interactions (this corresponds to the first stage of evolution of Cohen and Kaplan's *thermion*);
3. Once the temperature drops below T_D, as the Universe expands, the asymmetry can no longer change and is frozen. Then a net asymmetry remains:
$$\frac{n_B}{s} \approx \left. \frac{\dot{R}}{M_*^2 T} \right|_{T_D} . \tag{6.1.3}$$

Albeit Davoudiasl approach is close related to spontaneous baryogenesis, it poses several advantages. For instance, the *thermion* ϕ has to be added by hand, whereas the term in (6.1.2) is expected to be present in an effective theory of gravity. Moreover, the initial conditions for ϕ need to be specified and reasoned: it must be forced to evolve homogeneously, in the beginning, and must be spatially uniform. In contrast, the time evolution of $R \sim H^2$ is required in a cosmological background and it is highly spatial uniform because the Universe is highly homogeneous.

Note that, in spite of the "gravitational" attribute, gravity itself does not play any special role in this type of model for baryogenesis: it is only the background with which the baryon current interacts. Hence, this mechanism develops as spontaneous baryogenesis, but in a gravitational background — instead of motivated by the motion of an *ad hoc* scalar field.

In GR, $R = -8\pi GT = -8\pi G(1-3\omega)\rho$, where $\omega = p/\rho$ is the equation of state (EOS) parameter and T is the trace of the energy-momentum tensor. As such, in the radiation dominated epoch $\omega = 1/3$ so that $R = \dot{R}$ vanishes and no net baryon number asymmetry can be generated. Therefore, to develop gravitational baryogenesis, some modification must be introduced.

In his work (2004), Davoudiasl seeks deviations from $\omega = 1/3$. He shows that it is possible to obtain the correct magnitude for the B-asymmetry in many different cosmological scenarios [54]: for example, the numerical value of η can be obtained for $\omega \approx 1/3$, resorting to trace anomalies that occur in the context of Particle Physics and choosing the appropriate mass scales. Looking at equation (6.1.3), though, one thing that stands out is the fact that one only gets a non-vanishing η for a non-vanishing $\dot{R}$. This is zero in GR, so we are naturally led to modified gravity theories, as the ensuing modified equations of motion can provide very different relations between R and T and evade this limitation.

Due to its recent successes in explaining cosmological mysteries, as pointed out in the last chapter, many authors have resorted to $f(R)$-theories to study baryogenesis. In this context, gravitational baryogenesis may occur, provided the form of the function $f(R)$ is nearly linear in R [55].

The purpose of the present chapter is to investigate how a NMC can impact gravitational baryogenesis and determine how the correct value for the baryon asymmetry constrains its parameter space, thus extending the previous work. The following sections correspond to the original results obtained in Ref. [2].

Among other features, the NMC gives rise to the non-conservation of the energy-momentum tensor (5.3.32). Following the equivalence with a two-scalar field model, this may be recast as an energy exchange between matter and the former. One of these scalar fields is dynamically identified with the scalar curvature (as obtained in section 5.3.2), so that the transition from spontaneous to gravitational baryogenesis appears quite naturally.

6.2 The Model

As usual, we use the statistical results we have found before — $n_B = \frac{g_b}{6}\mu_B T^2$ and $s = \frac{2\pi^2}{45} g_{*S} T^3$, for the net particle number and entropy density, respectively — to evaluate the baryon-to-entropy ratio:

$$\eta_S \equiv \frac{n_B}{s} \approx -\frac{15 g_b}{4\pi^2 g_*}\frac{\dot{R}}{M_*^2 T} \ , \quad \text{at } T = T_D \ , \tag{6.2.4}$$

considering the number of intrinsic degrees of freedom of baryons to be $g_b \sim O(1)$ and $g_{*S} \approx g_* \sim 107$, at this early epoch.

We consider a flat Universe with a FRW metric

$$ds^2 = dt^2 - a^2(t) dV^2 \ , \tag{6.2.5}$$

and matter is assumed to behave as a perfect fluid, with an energy-momentum tensor given by equation (3.1.3). Then, the Ricci scalar curvature is defined as

$$R = -6\left[\left(\frac{\dot{a}}{a}\right)^2 + \frac{\ddot{a}}{a}\right] \ , \tag{6.2.6}$$

and allows us to rewrite the field equations in terms of the scale factor. The tt component of equations (5.3.28) yields the Modified Friedmann equation:

$$-3\frac{\ddot{a}}{a}F = \frac{1}{2}f_2\rho - 3\frac{\dot{a}}{a}\dot{F} + \frac{1}{2}\kappa f_1 \ . \tag{6.2.7}$$

Likewise, the rr component of the field equations reads:

$$\left(2\frac{\dot{a}^2}{a^2}+\frac{\ddot{a}}{a}\right)F=\frac{1}{2}f_2p+\ddot{F}+2\frac{\dot{a}}{a}\dot{F}-\frac{1}{2}\kappa f_1 \ , \tag{6.2.8}$$

while the trace corresponds to

$$3\left(\ddot{F}+3\frac{\dot{a}}{a}\dot{F}\right)-6\left(\frac{\dot{a}^2}{a^2}+\frac{\ddot{a}}{a}\right)F=\frac{1}{2}f_2(3p-\rho)+2\kappa f_1 \ . \tag{6.2.9}$$

Given that the NMC gives rise to an explicit dependence of the field equations on the Lagrangian density, we recall that radiation is not only composed of photons, but also of relativistic particles. Following the discussion of section 5.3.3, we adopt the form $\mathcal{L}_f=\rho_B$ for the Lagrangian density of the latter, while for photons we have $\mathcal{L}_\gamma=-p_\gamma$ (notice the sign change due to the adopted metric signature).

Since the energy and number densities and the pressure are related through $p=n(\partial\rho/\partial n)-\rho$, the former scales as $\rho\sim n^{1+\omega}$, so that both relativistic baryons as well as photons obey $\rho\sim n^{4/3}$, enabling

$$\rho=\rho_B+\rho_\gamma=\rho_\gamma\left(1+\frac{\rho_B}{\rho_\gamma}\right)=\rho_\gamma\left[1+\left(\frac{n_B}{n_\gamma}\right)^{4/3}\right]=\rho_\gamma\left(1+\eta^{4/3}\right) \ . \tag{6.2.10}$$

By the same token, the total Lagrangian is the sum of the Lagrangian densities of each species:

$$\mathcal{L}=\rho_B-p_\gamma=(\rho-\rho_\gamma)-p_\gamma=\rho-\frac{4}{3}\rho_\gamma=\rho\left[1-\frac{4}{3}\left(\frac{1}{1+\eta^{4/3}}\right)\right] \ . \tag{6.2.11}$$

Now that we have written the Lagrangian explicitly, equation (5.3.32) becomes:

$$\dot{\rho}+4\frac{\dot{a}}{a}\rho=\frac{f_2'(R)}{f_2(R)}\left[\rho_B-p_\gamma-(\rho_B+\rho_\gamma)\right]\dot{R}=-\frac{4}{3}\frac{f_2'(R)}{f_2(R)}\rho_\gamma\dot{R}=-\frac{4}{3}\frac{f_2'(R)}{f_2(R)}\frac{\rho}{1+\eta^{4/3}}\dot{R} \ . \tag{6.2.12}$$

This can be directly integrated, considering $\eta=$ const., which is a good approximation if we neglect particle-antiparticle annihilation (below T_D) and ignore other processes, such as the production of photons in stars, as the majority is absorbed by nearby objects:

$$\rho(t)=\rho_0 f_2(R(t))^{-\frac{4}{3(1+\eta^{4/3})}}a(t)^{-4} \ , \tag{6.2.13}$$

where ρ_0 is the the energy density at an arbitrary time $t=t_0$. In the absence of a NMC, we recover the usual dependence $\rho\sim a^{-4}$ for a radiation dominated Universe.

6.3 Cosmological Dynamics

To determine the cosmological dynamics depicted in the previous section, we now make the *Ansatz* that the scale factor evolves as a power-law, $a(t)\sim t^\alpha$ (with $\alpha>0$), so that

$$H(t)=\frac{\alpha}{t} \ , \quad R(t)=6\frac{\alpha(1-2\alpha)}{t^2} \ . \tag{6.3.14}$$

We also adopt power-law forms for the functions present in the action functional (5.3.20),

$$f_1(R) = R\left(\frac{|R|}{M_1^2}\right)^m \ , \tag{6.3.15}$$

$$f_2(R) = \left(\frac{|R|}{M_2^2}\right)^n \ , \tag{6.3.16}$$

where M_i are characteristic mass scales; both m and n should be close to zero, in order to seek only slight deviations from GR. Also, $1+m$ and n must be greater than zero, so that no divergences in the action functional occur.

We take the absolute value of the scalar curvature to allow it to be negative (*i.e.* $\alpha > 1/2$); alternatively, one could have considered negative values for M_i^2, although such notation is less appealing. Notice that, if we had adopted the metric signature $(-1,1,1,1)$, the curvature would change sign and we would be excluding the reciprocal region $0 < \alpha < 1/2$ — effectively attributing physical significance to the metric signature. This detail was overlooked in Ref. [55], which as a result only studied half of the allowed parameter space.

Following equation (6.2.13), we have[1]:

$$\rho(t) = \rho_0 \left(\frac{t}{t_0}\right)^{4\left(\frac{2}{3}\frac{n}{1+\eta^{4/3}}-\alpha\right)} . \tag{6.3.17}$$

Replacing equations (6.3.14) and (6.3.17) into the modified field ones (6.2.7-6.2.9), we may obtain a relation between the exponents α and n. Considering (6.1.1), we neglect the extremely small value of the baryon-to-photon ratio, obtaining

$$\alpha = \frac{1}{2}\left(1+m+\frac{n}{3}\frac{1-3\eta^{4/3}}{1+\eta^{4/3}}\right) \approx \frac{1}{2}\left(1+m+\frac{n}{3}\right) . \tag{6.3.18}$$

Substituting into equations (6.2.7) or (6.2.8) and solving for ρ_0, we obtain the energy density

$$\rho(t) = h_{mn} M_P^4 \left(\frac{M_P}{M_1}\right)^{2m} \left(\frac{M_2}{M_P}\right)^{2n} (M_P t)^{2(n-m-1)} , \tag{6.3.19}$$

defining the dimensionless quantity

$$h_{mn} \equiv \frac{(3m+n)(3+3m+n)[3+n-m(6+15m+n)]}{2[3m(6+5n)+n(3+n)]} \times \left[\left(1+m+\frac{n}{3}\right)|3m+n|\right]^{m-n} . \tag{6.3.20}$$

6.3.1 Baryon Asymmetry

We now resort to the usual result for the energy density (3.2.43), to write the temperature as a function of the latter,

$$T = M_P \left(\frac{30}{g_* \pi^2} h_{mn}\right)^{1/4} \left(\frac{M_P}{M_1}\right)^{m/2} \left(\frac{M_2}{M_P}\right)^{n/2} (M_P t)^{(n-m-1)/2} . \tag{6.3.21}$$

From definition (6.2.6), it follows that

$$\dot{R} = \frac{12(2\alpha-1)\alpha}{t^3} ; \tag{6.3.22}$$

and the net baryon asymmetry can be written as

$$\eta_S \approx \frac{g_b}{g_*}\frac{45}{\pi^2}\frac{\alpha(2\alpha-1)}{t_D^3 T_D M_*^2} , \tag{6.3.23}$$

[1]We still need to verify if all these power-laws are compatible with the field equations. Moving all the terms into the r.h.s., we end up with three equations in the form $0 = \mathrm{Eq}_i$ with $i = 1,2,3$ corresponding to (6.2.7), (6.2.8) and (6.2.9), respectively. The trace of the stress-energy tensor for a perfect fluid $T = \rho - 3p$ vanishes if we are dealing with relativistic matter. Then, the compatibility test consists on assuring that $\mathrm{Eq}_3 - (\mathrm{Eq}_1 - 3\mathrm{Eq}_2) = 0$, which is easily proven to be true.

where t_D is the decoupling time, at which the baryon violating interactions go out of equilibrium. Notice that the above can become negative, signalling the excess production of antimatter[2]: this could be corrected by changing the sign of the interaction term (6.1.2).

Inverting equation (6.3.21), we obtain the relation $t = t(T)$, which we insert into equation (6.3.23) to obtain

$$\eta_S \lesssim \frac{5}{2\pi^2}\frac{g_b}{g_*}h'_{mn}\left(\frac{M_P}{M_*}\right)^2\left(\frac{T_D}{M_P}\right)^{\frac{5-m+n}{1+m-n}}\left[\pi\sqrt{\frac{g_*}{30}}\left(\frac{M_1}{M_P}\right)^m\left(\frac{M_2}{M_P}\right)^{-n}\right]^{3/(1+m-n)} , \tag{6.3.24}$$

where we have defined the dimensionless quantity

$$h'_{mn} \equiv (3+3m+n)(3m+n)(h_{mn})^{3/2(n-m-1)} . \tag{6.3.25}$$

As pointed out previously, it is natural to expect an operator such as (6.1.2) in the low effective field theory, if the cutoff scale M_* is of the order the reduced Planck mass M_P. For this choice of M_*, the baryon asymmetry generated can be sufficiently large for $T_D = M_I$ [54], where $M_I \approx 2\times 10^{16}$ GeV is the upper bound on the energy scale of inflation, as placed by the Wilkinson Microwave Anisotropy Probe (WMAP) three-year data set [56]. It is crucial that T_D is placed after inflation, so that the asymmetry fixed once B-violating interactions decouple is not diluted by the ensuing exponential growth[3].

As hinted from the expression above, η_S is very sensitive to the inflationary energy scale $M_I \sim T_D$: therefore, it is relevant to study numerically how the constraints on the exponents (n,m) and mass scales M_i are affected by the choice of the decoupling temperature T_D.

6.3.2 Big Bang Nucleosynthesis

We now assess how the gravitational mechanism detailed in the previous sections can generate the correct amount of baryon asymmetry while maintaining compatibility with the typical temperatures $\sim 0.1-100$ MeV of BBN, the next major phase in the early Universe. Standard Cosmology sets the starting point of BBN very close to $T \approx 1$ MeV, when the weak interactions freeze-out; in our framework, this can be extended to a higher value, due to the modified Hubble parameter that is used to define the Universe expansion rate; and to a lower one — until about 0.1 MeV — which is characteristic of the temperature at which the mass fractions of the primordial elements get close to unity. This serves to justify the range of typical temperatures that we allow, taking into consideration the possible conditions that can delay the production of the abundances, like properties of the elements themselves (such as their binding energy) or the usual "bottlenecks", the fact that the lack of light elements can prevent the production of heavier

[2] In fact, it makes no difference to say that, instead of matter, everything is made up of antimatter; the crucial point is that some of the two had to gain dimension over the other, in order to explain the existence of structures in the Universe, instead of their annihilation.

[3] Actually, to guarantee this is the case, T_D should be placed below M_I. However, as the inflationary scale is the only observational bound we can use to constrain the present model, we consider the limiting case $T_D = M_I$ to explore the general features of the latter.

ones (since they participate in the formation reactions). All of these questions were studied in section 3.5 of the present work.

Baryogenesis must strictly occur before BBN, so that the initial conditions are in place to build up the observed abundances that were produced in the early time, when the energy and number density were dominated by relativistic particles. At this stage of the evolution of the Universe, protons and neutrons are kept in thermal equilibrium by weak interactions, due to their rapid collisions. The weak interaction rate $\Lambda(T)$ is determined by means of the conversion rates of protons into neutrons. It corresponds to the sum of the decay rates of each reaction in (3.5.83), plus the inverse ones. At sufficiently high temperatures, it is given by (3.5.97):

$$\Lambda(T) \approx \frac{7\pi}{60}(1+3g_A^2)G_F^2 T^5 \ , \tag{6.3.26}$$

where $G_F \approx 1.166\times10^{-5}$ GeV^{-2} is the Fermi coupling constant and $g_A \approx 1.27$ is the axial-vector coupling constant of the nucleon.

We are interested in the freeze-out temperature, T_f, at which the baryons decouple from leptons. To compute it, one has to equate the rate of the weak interactions to the expansion rate of the Universe, $\Lambda(T) \sim H$, since — from then on — the weak interaction rates are comparatively slower and the primordial abundances start being produced.

Using $g_* = g_*^{\mathrm{BBN}} = 10.75$ and equations (6.3.14) and (6.3.21), it follows that

$$T_f = \left[\frac{1}{\pi}\sqrt{\frac{30 h_{mn}}{g_*^{\mathrm{BBN}}}}\left(\frac{7\pi(1+3g_A^2)}{10(3+3m+n)}G_F^2 M_P^4\right)^{1+m-n}\left(\frac{M_P}{M_1}\right)^m\left(\frac{M_2}{M_P}\right)^n\right]^{1/(5n-5m-3)} M_P \ . \tag{6.3.27}$$

6.4 Parameter Constraints

Given the results obtained above, we impose the following set of requirements for the allowed values of the exponents (n, m):

- The density, given by equation (6.3.19), must be positive defined;
- We consider only small deviations from GR, $m \sim 0$ and $n \gtrsim 0$;
- An expanding Universe requires that $\alpha > 0$, so that $m > -(1 + n/3)$.

Using equation (6.3.19), this yields,

- For $\alpha > 1/2$, the condition

$$-\frac{(3+n)n}{18+15n} < m < \frac{\sqrt{216+72n+n^2}-n}{30} - \frac{1}{5} \approx 0.24 \ , \tag{6.4.28}$$

- For $\alpha < 1/2$, m can only take negative values,

$$-0.69 \approx -\frac{\sqrt{216+72n+n^2}+n}{30} - \frac{1}{5} < m < -\frac{n}{3} \ . \tag{6.4.29}$$

Compatibility between Baryogenesis and BBN

We now ascertain what are the allowed values for the exponents (m, n) and mass scales M_i compatible with the observed amount of asymmetry $\eta_S^{\rm obs}$ and with a freeze-out temperature in the range $[0.1, 100]$ MeV. This allows us to use the same gravity functions, $f_1(R)$ and $f_2(R)$, for explaining the dynamics of two (sequential) major phases of the thermal history of our Universe.

To do so, we equal η_S to its observational value and solve equation (6.3.24) for the combination $(M_P/M_1)^m\,(M_2/M_P)^n$. We replace this into (6.3.27), to finally obtain

$$T_f = T_D\left[\sqrt{\frac{g_*^{\rm BBN}}{g_*^{\rm Bar}}}\left[\frac{400}{343\pi}\,\frac{\eta_S^{\rm obs}}{(1+3g_A^2)^3}\,\frac{g_*^{\rm Bar}}{g_b}\,\frac{(3+3m+n)^2}{3m+n}\left(\frac{M_*}{G_F^3T_D^7}\right)^2\right]^{\frac{1+m-n}{3}}\right]^{1/(3+5m-5n)}, \qquad (6.4.30)$$

where $g_*^{\rm Bar}\approx 107$ corresponds to the relativistic degrees of freedom of species at $T = T_D \sim 10^{16}$ GeV, when the full set of Standard Model particles is effectively massless.

Imposing the requirements outlined at the beginning of this section and the constraint $10^{-4} < T_f < 10^{-1}$ GeV yields the allowed combinations of exponents (n, m) shown in figure 6.1, for different choices of the decoupling temperature T_D: as can be seen, the latter does not impact strongly on the overall shape of the allowed region — indeed, a smaller T_D slightly shifts the allowed region into the lower right corner of the (n, m) plane.

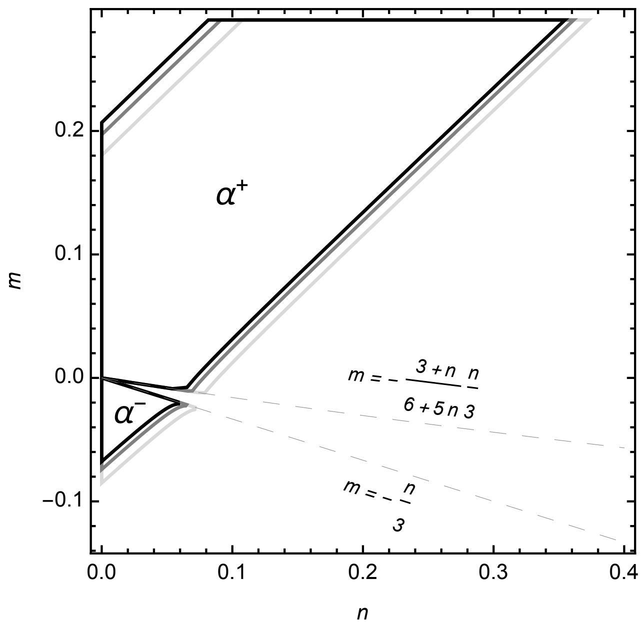


Figure 6.1: Allowed regions for the exponents (n, m): α^- and α^+ correspond to $0 < \alpha < 1/2$ and $\alpha > 1/2$, respectively. From lighter to darker shade, $T_D = \{1, 2, 3\}\times 10^{16}$ GeV.

In the past decade, the upper bounds on the inflation mass scale were refined from 3.3 [57] to 2×10^{16} GeV [56]. Although this has only a minor impact on the allowed region for the exponents (n, m), it significantly alters the mass scale M_i of our model, as we will discuss next.

Admitting $T_D = 2 \times 10^{16}$ GeV and considering a trivial NMC ($n = 0$), we conclude that all values between $-0.07 \lesssim m \lesssim 0.19$ are allowed, although a more precise measurement of M_I could lower the upper limit of this range. Conversely, if we isolate the effect of the NMC (setting $m = 0$), we find that any value of its exponent in the interval $0 < n \lesssim 0.07$ is allowed.

As an example of how the decoupling temperature scale significantly alters the mass scales M_i of our model, figure 6.3 presents the scenario where $M_1 = M_2 = M_P$, adopting the old bound for M_I as the value of the decoupling temperature, $T_D = 3.3 \times 10^{16}$ GeV, showing that the right amount of baryon asymmetry $\eta_S \sim 10^{-10}$ can be attained.

However, adopting the current bound for the inflationary energy scale, so that $T_D = 2 \times 10^{16}$ GeV, figure 6.2 shows that the ensuing baryon asymmetry is insufficient, $\eta_S \lesssim 7 \times 10^{-12}$, thus disallowing the possibility of a single mass scale for our model, $M_1 = M_2$, to be of the order of Planck's scale.

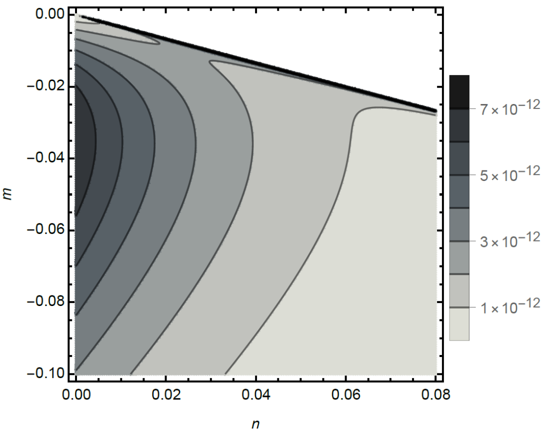


Figure 6.2: Baryon-to-entropy ratio η_S (6.3.24) contour plot for the particular choice $M_1 = M_2 = M_P$ and $T_D = 2 \times 10^{16}$ GeV.

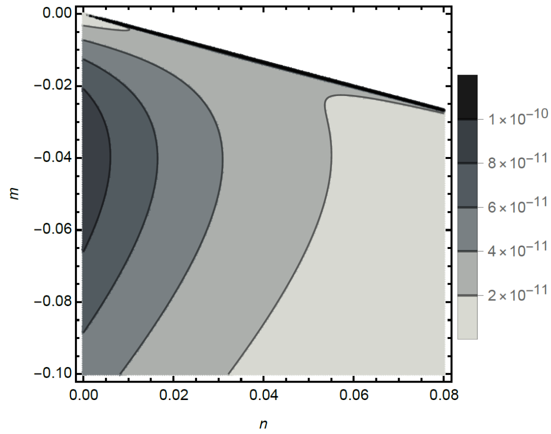


Figure 6.3: Same as figure 6.2, but with $T_D = 3.3 \times 10^{16}$ GeV.

6.5 Baryogenesis and $f(R)$ theories

In this section, we consider a minimal coupling $n = 0$, aiming at generalizing the results obtained in Ref. [55] for the exponent m and the mass scale M_1.

We solve equation (6.3.24) for M_1 with $n = 0$, to obtain it as a function of the exponent m, with a dependence on the inflationary and Planck mass scales of the form

$$M_1^{3m} \sim M_P^3 M_*^{2(1+m)} T_D^{m-5} \,. \tag{6.5.31}$$

Since $m \sim 0$, this mass scale turns out to be very sensitive to the inflationary mass scale $M_I \sim T_D$, as discussed below.

Figure 6.4 shows this behaviour for $T_D = 3.3 \times 10^{16}$ GeV. The maximum value $M_1 \approx 1.5 M_P$ is attained for $m = -0.04$. Drastically smaller mass scales are obtained for the current constraint

$T_D = 2 \times 10^{16}$ GeV, as graphed in figure 6.5: M_1 is no longer of the order of M_P, but instead six orders of magnitude below, $M_1 \sim 10^{12}$ GeV.

This signals the strong dependence of the mass scale M_1 on the decoupling temperature T_D for small values of the exponent m, as depicted in figure 6.6 and exemplified in table 6.1: indeed, equation (6.5.31) shows that for $m \sim 0$, $M_1 \sim T_D^{-5/3m}$; only for the unphysical case of very large deviations from GR is this alleviated, since $|m| \gg 1$ implies that $M_1 \sim T_D^{1/3}$.

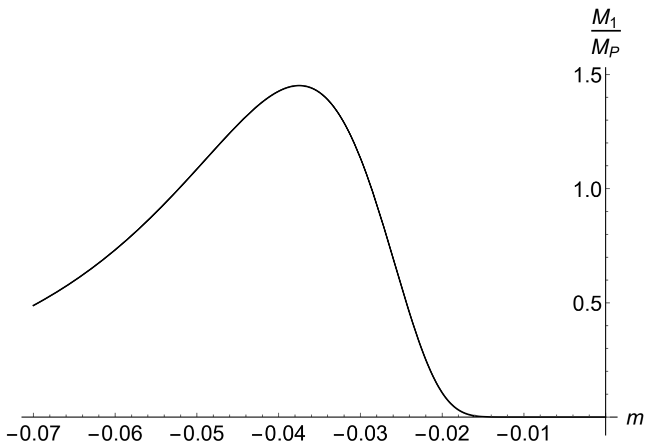


Figure 6.4: Plot of M_1 vs m for $T_D = 3.3 \times 10^{16}$ GeV.

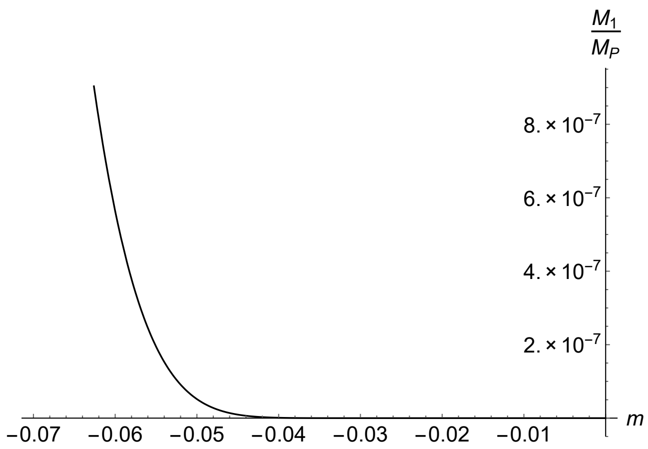


Figure 6.5: Plot of M_1 vs m for $T_D = 2 \times 10^{16}$ GeV.

6.6 Gravitational Baryogenesis with a NMC

A new result of our work is that a NMC can, by itself, induce gravitational baryogenesis. In order to isolate its effect, we set $m = 0$ in the previous expressions, and conclude that the allowed values for n compatible with BBN and the observed B-asymmetry lie within the interval $0 < n \lesssim 0.07$ (plot 6.1).

Following the same argument as in the previous paragraph, we now consider both functions (6.3.24)

$f_1(R)$	T_D (GeV)	M_1 (GeV)
$\sim R^{0.99}$	3.3×10^{16}	2×10^{10}
	2×10^{16}	8×10^{-27}
$\sim R^{0.96}$	3.3×10^{16}	3×10^{18}
	2×10^{16}	3×10^{9}
$\sim R^{0.93}$	3.3×10^{16}	1×10^{18}
	2×10^{16}	8×10^{12}

Table 6.1: Dependence of M_1 on the exponent m for different choices of the decoupling temperature T_D.

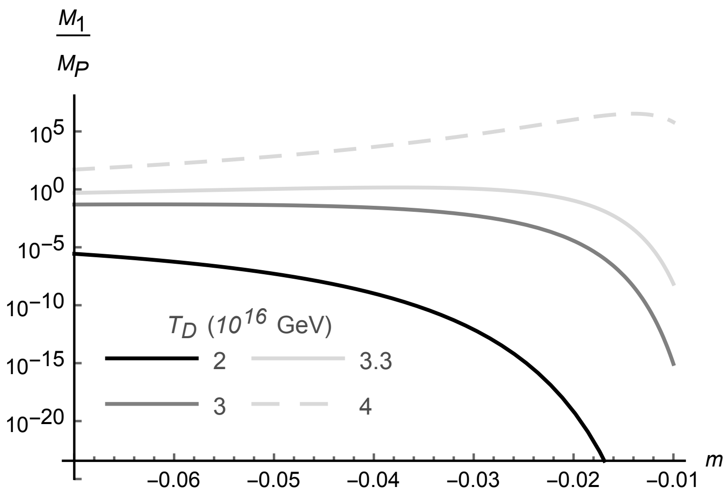


Figure 6.6: Dependence of M_1 on the exponent m for different choices of the decoupling temperature.

and (6.3.27) with $m = 0$, in the regime where $\alpha > 1/2$, and solve them for the characteristic mass scale M_2: this yields the scaling law

$$M_2^{3n} \sim M_P^{-3} M_*^{2(n-1)} T_D^{5+n} , \tag{6.6.32}$$

which, since $n \sim 0$, also leads to the conclusion that the mass scale of the NMC is very sensitive to the value adopted for the decoupling temperature, $M_2 \sim T_D^{\frac{5}{3n}}$. This is clearly shown in figure 6.7: we get a large discordance between the mass functions for small deviations from GR, as in the previous case. On the contrary, for $n \gg 1$, the mass scale shows a less pronounced dependence on the decoupling temperature: $M_2 \sim T_D^{1/3}$.

In order to restrict, with detail, the allowed range for the NMC exponent, we can solve numerically the intersection of the functions (6.3.24) and (6.3.27), equating $\eta_S = \eta_S^{\rm obs}$ and considering the limiting temperature $T_f \approx 10^{-4}$ GeV, characterizing BBN. The results are presented in table 6.2: different choices of T_D have a major effect on the mass scale, altering only minimally (at the third decimal place) the function exponent.

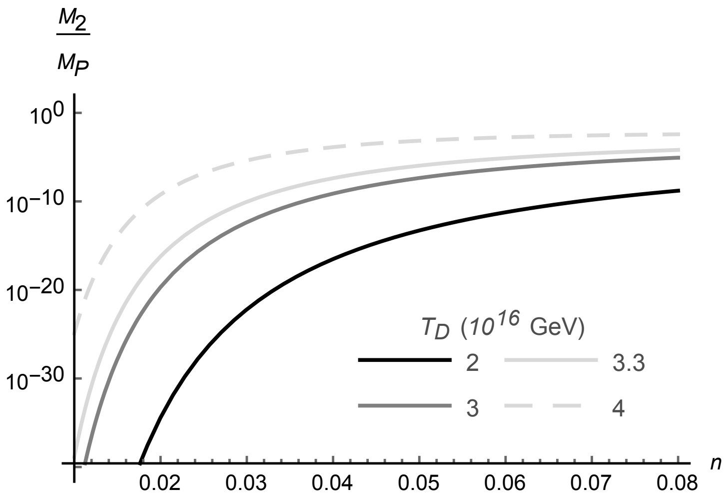


Figure 6.7: Dependence of M_2 on the exponent n for different choices of the decoupling temperature.

$f_2(R)$	T_D (GeV)	M_2 (GeV)
$\sim R^{0.070}$	3.3×10^{16}	6.5×10^{13}
$\sim R^{0.078}$	2.0×10^{16}	2.5×10^{9}
$\sim R^{0.089}$	1.0×10^{16}	3.5×10^{4}

Table 6.2: Mass scale M_2, as determined by the intersection of function (6.3.24) with (6.3.27), depending on the choice of T_D.

6.7 Entropy Conservation

As the baryon number to entropy ratio is paramount to our study, it is relevant to assess how the non-conservation law for the energy-momentum tensor may affect adiabaticity.

In standard cosmology, the total entropy does not change as the Universe expands: since we know that — at low energies — there are no decays in which baryon number is created or destroyed, the baryon-to-entropy ratio η_S remains a constant. Similarly, once large scale annihilation processes have ended, the baryon-to-photon ratio η is also constant, and both quantities can be swiftly related.

To assess the impact of a NMC, we resort to the first law of thermodynamics,

$$TdS = dE + pdV \ , \tag{6.7.33}$$

where $E = \rho(aL)^3$ and $S = s(aL)^3$ are the internal energy and entropy contained in an arbitrary comoving volume of size L, respectively, so that

$$TdS = d(\rho a^3) + pd(a^3) \rightarrow \frac{T}{a^3}\dot{S} = \dot{\rho} + 4H\rho \ . \tag{6.7.34}$$

Using equations (3.2.51), (6.2.12) and (3.2.43) leads to

$$(1+\eta^{4/3})\frac{g_{*S}(T)}{g_*}\frac{\dot{S}}{S} = -\frac{f_2'}{f_2}\dot{R} \ . \tag{6.7.35}$$

Assuming the baryon-to-photon ratio to be approximately constant, as previously discussed, and taking $g_{*S}(T) \sim g_*$ allows us to directly integrate the above, obtaining

$$S(t) \sim f_2(R(t))^{-\frac{1}{1+\eta^{4/3}}} \approx f_2(R(t))^{-1} \ . \tag{6.7.36}$$

so that the entropy remains constant in the absence of a NMC. Its variation can be neglected if it occurs at a rate much smaller than the expansion rate of the Universe,

$$\left|\frac{\dot{S}}{S}\right| \approx \left|\frac{f_2'}{f_2}\dot{R}\right| \ll H \ . \tag{6.7.37}$$

Inserting equations (6.3.18), (6.3.14) and (6.3.15) yields the simple condition $|(11/3)n - m| \ll 1$, which is naturally satisfied for the small perturbations $n, m \sim 0$ considered in the preceding section. Thus, we are led to conclude that the entropy remains approximately constant during gravitational baryogenesis, so that $\eta \sim \eta_S \approx$ const..

6.7.1 Entropy for Irreversible Matter Creation

We have interpreted the r.h.s. of equation (5.3.32) as an *extra* thermodynamic flow that increases the energy of particles. Equivalently, this could by interpreted as leading to irreversible matter creation, following the discussion of section 5.3.4.

Following that line of reasoning, the presence of the NMC gives rise to a heat exchange between matter and gravitational fields, which allows us to compute the timescale for entropy which can account for the production of particles in the early Universe.

The framework of open systems gives a physical meaning to the heat flow appearing in the presence of the NMC, instead of just posing its existence. Unlike the general interpretation of the last section, this framework identifies an origin for the additional entropy based on the irreversible flow from the gravitational to matter fields.

For our discussion, it suffices to show that, in spite of these exchanges in the open *matter-gravitation* system, the Universe can be considered nearly isentropic during its evolution. Otherwise, although the baryon asymmetry is fixed after the B-violating forces decouple, the number of baryons could vary in an expanding Universe.

In section 5.3.4, we have already obtained the timescale for entropy production:

$$\frac{1}{S}\frac{dS}{dt} = \frac{1}{\rho+p}\frac{d[\ln f_2(R)]}{dt}(\mathcal{L}-\rho) \geq 0 \ . \tag{6.7.38}$$

Using the Lagrangian density (6.2.11), with $\eta \lll 1$, this can be simplified to:

$$\frac{1}{S}\frac{dS}{dt} = -\frac{d[\ln f_2(R)]}{dt} = \frac{2n}{t} \ , \tag{6.7.39}$$

which turns out to be the same result as the last section, consistent with a negligible variation of entropy and thus justifying the use of the observed η_S, to characterize the cosmological evolution below T_D. Note also that this framework does not allow negative values for the exponent of the NMC function.

Chapter 7

Conclusions

7.1 Overview

In this work, we have studied how an extension of General Relativity — promoting the trivial terms in the Einstein-Hilbert action to generalized gravity functions — affects a mechanism for the generation of the baryon asymmetry. This is possible via an effective coupling between the net baryonic current and the derivative of the Ricci scalar, that dynamically breaks CPT invariance.

In the context of the Standard Model Extension, Lorentz-violating terms arise as expectation values of Lorentz tensors that could undergo through spontaneous symmetry breaking, in the context of string theories. This can lead to spontaneous CPT breaking too. Identifying CPT with the strong reflection operation, we could quickly identify all the CPT-even or odd terms that can be constructed in this type of phenomenological models. As discussed, a class of CPT-odd terms, originating in this bodywork, can give rise to the right amount of baryon asymmetry, as already studied in Ref. [17].

Although we followed a fundamentally different path, the previous model is based on the violation of CPT invariance, allowed by a background field: this is the same idea of our work. However, we cannot interpret $\langle \partial_\mu R \rangle$ as a SME background, that is, a field gaining an expectation value that permeates space. Instead, $\dot{R}$ is the natural quantity to gauge the evolution of the cosmological background, that exists due to the structure of spacetime itself and which cannot be altered in a local experiment.

Moreover, since it becomes smaller as the temperature drops, the interaction term $(\partial_\mu R) J^\mu$ becomes ineffective as the Universe expands, as required by the CPT invariant ground state of our low-energy Standard Model. It is then the standard cosmological evolution that determines the size of the dynamical breakdown of CPT invariance.

We have learnt that particle and antiparticle thermal distributions are the same, in the primordial Universe, since the moment CPT invariance is established (as the CPT Theorem predicts the equality of masses and decay widths). Reversely, if an effective interaction term originates an energy shift for this pair, they are allowed to equilibrate with different thermal weights. Therefore, in the period when CPT is temporarily broken, we can promote baryogenesis in a thermal equilibrium scenario, as this energy shift is equivalent to a chemical potential — providing it is slow-varying enough —, allowing us to compute the baryon-to-entropy ratio. In particular, we studied how the inclusion of a NMC can affect an isentropic expansion, which is fundamental to justify the mapping of the present η parameter to the value it read once the asymmetry was fixed. The results of this work were reported in Ref. [2].

Considering the non-conservation of the energy-momentum tensor, we integrated the First Law of Thermodynamics to read the evolution of the entropy: albeit it is not constant as in GR, we asserted that the time scale on which it varies significantly is much larger than the Hubble time, as long as the constraint $|(11/3)n - m| \ll 1$ is kept: since small deviations from GR imply very small exponents n and m, this is trivially fulfilled for all cases approached.

We have constrained the parameter space of a model with both a non-linear curvature term and a NMC: we showed that the observed amount of baryon asymmetry is attained with only small deviations from GR, while keeping compatibility with the typical temperatures of Big Bang Nucleosynthesis. We also concluded that the characteristic mass scales are very sensitive to the value for the decoupling temperature T_D, at which the baryon violation interactions go out of equilibrium and the baryon to entropy density becomes fixed — which we admit to be of the same order of magnitude as the energy scale of inflation.

We have also extended the parameter space of Ref. [55] by allowing for both positive and negative values of the Ricci scalar curvature, as its sign changes with the adopted metric signature and has no physical significance: we showed that a curvature term of the form $f_1(R) \sim R^{1+m}$ with $-0.07 \lesssim m \lesssim 0.19$ can, by itself, generate the right amount of asymmetry. Although the allowed range of exponents is not very large (nor can it be, as we expect small deviations from GR, $m \sim 0$), the characteristic mass scale M_1 can vary significantly — again depending crucially on the choice for the free cosmological parameter T_D.

Finally, we found that a NMC of the form $f_2(R) \sim R^n$ is consistent with the observed η-parameter and BBN, with a small exponent in the range $0 < n \lesssim 0.078$. As expected from the application to $f(R)$ theories, the characteristic mass scale M_2 is again highly sensitive to the value of the decoupling temperature.

Future work could focus on:

- the phenomenological consequences of the adopted form for the action functional, namely possible instabilities and the GR limit of the modified gravitational equations with the effective coupling present;

- extracting independent estimates on the characteristic mass scales M_1 and M_2, considering other cosmological phenomena occurring at the same temperature scale. These could then be used to better restrict the allowed exponents (n, m) and further assess what values of T_D lead to the desired amount of baryogenesis — and how these compare with the ever improving bounds on the energy scale of inflation.

7.2 Perspectives and Open Questions

The Standard Model — considering its construction, which requires renormalizable, Lorentz and gauge invariant terms in the Lagrangian density — explains why baryon and lepton numbers are conserved (up to a very good approximation). The baryon asymmetry, however, shows us that baryon number violation must occur in the fundamental laws. The question is what might be the scale associated with this violation.

As we have been arguing, Lorentz invariance and thus the CPT symmetry might not have been already established in the very early Universe. With this idea in mind, we considered a dimension six operator that temporarily violates CPT in the primitive Universe. We assumed, however, that particles were already in a thermal bath with the photons, at this stage. In our model, two major problems that could arise in *Planck scale baryogenesis* are thus avoidable. First, how can a small dimensionless quantity like η arise and be computed at this scale — this is solved because the departure from thermal equilibrium is no longer necessary and the standard thermal distributions can be used. Second, the rapidly expansion of the Universe during this period, which would dilute the baryon asymmetry by an enormous factor (at least e^{60}) — this does not occur in our model because the decoupling temperature is identified with the energy scale of inflation.

In spite of these considerations, this is highly speculative. Several other mechanisms have been proposed to understand the baryon asymmetry [58]. GUT models, for example, in which the phase transition associated with the breaking of the unified gauge group can lead the Higgs boson to acquire large masses and decay through $B-$violating interactions. The Standard Model satisfies itself all the elements for baryogenesis, when we bring instanton fields to our discussion: at sufficiently high temperatures, the barrier suppression can be overcome and a baryon asymmetry might be produced. This kind of processes can also process an existing lepton number, producing a net baryon number from it — what is generally called leptogenesis —, as the chiral anomaly allows the separate baryon B and lepton L numbers to be violated, but conserves $(B-L)$. Applied to the Standard Model, this model is called electroweak baryogenesis, which is very sensitive to the mass of the Higgs boson. Generalizations of this picture, in composite Higgs models, can enlarge CP-violation and make electroweak baryogenesis feasible [59]. In turn, if supersymmetry is discovered, the so-called Affleck-Dine mechanism might appear quite plausible: without entering into detail, as fermions are thought to have boson partners that carry baryon and lepton numbers, the latter can decay into ordinary quarks in a phase transition during the early Universe, leaving space with a residual net baryon number.

In summary, the origin of the cosmic matter-antimatter asymmetry remains unknown. By now, we can address different problems in each approach, but one cannot establish that one or the other is correct.

We hope that future experimental tests, both cosmological and particle related, will be able to rule out some of these scenarios. In fact, cosmological CPT violation and its implications in baryo/leptogenesis models — via a coupling like $(\partial_\mu \phi J^\mu)$ — have been studied, leading us to believe that future measurements on the CMB polarization might detect this particular signature [60]. Even if this model is favored, one cannot exclude baryon and lepton number violating interactions which might be significant at scales well below the Planck scale.

More than describing the most promising possibility for baryogenesis, the purpose of this work was to explore the rich connection between different fields that this problem implies, from all our "a priori" statements — symmetry related — to our Universe's evolution history.

Choosing this specific idea, we studied — in detail — the CPT invariance (as well as the formal definition of each discrete symmetry), the evidence for a baryon asymmetry (excluding arguments of statistical fluctuations or an overall symmetric Universe) and the subsequent thermal phase of cosmological evolution, Big Bang Nucleosynthesis, which is strictly restricted by the former. Bringing modified gravity to this problem, we discussed how the non-minimal curvature-matter coupling yields interesting features to our discussion, like the explicit form of the Lagrangian in the field equations, whose on-shell degeneracy no longer holds; and the non-conservation of the energy-momentum tensor, which modifies the usual adiabatic energy conservation law, enabling us to interpret the NMC as an exchange of energy between matter and curvature.

We reviewed a recent thermodynamic interpretation of this non-minimal coupling which, in spite of being plausible [61] — consisting on a simple *extrapolation* of the Fundamental Laws of thermodynamics —, is not based on first principles when applied to the Universe itself (and its behaviour as any ordinary macroscopic thermodynamic system that tends to a maximum entropy state). The cosmological entropy evolution and, in particular, the thermodynamics of matter creation have been topics of interest in the literature [62,63].

To conclude, this work has opened my mind towards the necessity to develop a consistent formulation of thermodynamics in the context of the cosmological system, in the presence of a non-minimal geometry-matter coupling; and to the applicability of the discrete symmetries in an evolutionary Universe which tends to an isotropic and homogeneous state, in equilibrium. A lot of interesting problems arise in this framework. For example, how does the cosmological expansion provide an arrow of time and whether the time reversal violation in weak interactions is correlated with the latter. Another intriguing question is the following: if the Second Law of Thermodynamics applies to the Universe directly, are closed time curves still a possible solution?

Bibliography

[1] Antonio Riotto. Theories of baryogenesis. In *Proceedings, Summer School in High-energy physics and cosmology: Trieste, Italy, June 29-July 17, 1998*, pages 326–436, 1998.

[2] M. P. L. P. Ramos and J. Páramos. Baryogenesis in Nonminimally Coupled $f(R)$ Theories. *Phys. Rev.*, D96(10):104024, 2017.

[3] Walter Greiner and Joachim Reinhardt. *Field Quantization.* Springer-Verlag Berlin Heidelberg, 1996.

[4] Gerhart Luders. Proof of the TCP theorem. *Annals Phys.*, 2:1–15, 1957. [Annals Phys.281,1004(2000)].

[5] J. J. Sakurai. *Invariance Principles And Elementary Particles.* Princeton University Press, 1964.

[6] G. Grawert, G. Lders, and H. Rollnik. The tcp theorem and its applications. *Fortschritte der Physik*, 7(6):291–328, 1959.

[7] M. S. Sozzi. *Discrete Symmetries and CP Violation: From Experiment to Theory.* Oxford University Press, 2008.

[8] Gordon Baym. *Lectures on Quantum Mechanics.* Westview Press.

[9] C. Patrignani et al. Review of Particle Physics. *Chin. Phys.*, C40(10):100001, 2016.

[10] Gerhart Luders and Bruno Zumino. Some Consequences of TCP-Invariance. *Phys. Rev.*, 106:385–386, 1957.

[11] S. Okubo. Decay of the Σ^+ Hyperon and its Antiparticle. *Physical Review*, 109:984–985, February 1958.

[12] A. D. Sakharov. Violation of CP Invariance, c Asymmetry, and Baryon Asymmetry of the Universe. *Pisma Zh. Eksp. Teor. Fiz.*, 5:32–35, 1967. [Usp. Fiz. Nauk161,61(1991)].

[13] V. Alan Kostelecky and Robertus Potting. CPT and strings. *Nucl. Phys.*, B359:545–570, 1991.

[14] V. Alan Kostelecky. Topics in Lorentz and CPT violation. In *The role of neutrinos, strings, gravity, and variable cosmological constant in elementary particle physics. Proceedings, 29th International Conference on high energy physics and cosmology, Orbis Scientiae, Coral Gables, USA, December 14-17, 2000*, pages 57–68, 2001.

[15] O. Bertolami, D. Colladay, V. Alan Kostelecky, and R. Potting. CPT, strings and baryogenesis. In *New worlds in astroparticle physics. Proceedings, International Workshop, Faro, Portugal, September 8-10, 1996*, pages 356–361, 1997.

[16] Don Colladay and V. Alan Kostelecky. CPT violation and the standard model. *Phys. Rev.*, D55:6760–6774, 1997.

[17] O. Bertolami, Don Colladay, V. Alan Kostelecky, and R. Potting. CPT violation and baryogenesis. *Phys. Lett.*, B395:178–183, 1997.

[18] Ralf Lehnert. CPT and Lorentz-symmetry breaking: A Review. *Frascati Phys. Ser.*, 43:131–154, 2007.

[19] Joshua Frieman, Michael Turner, and Dragan Huterer. Dark Energy and the Accelerating Universe. *Ann. Rev. Astron. Astrophys.*, 46:385–432, 2008.

[20] João G. Rosa. *Introduction To Cosmology: Lectures.* Universidade de Aveiro.

[21] E. W. Kolb and M. S. Turner. *The early universe.* Westview Press, 1990.

[22] Rolf Kappl and Martin Wolfgang Winkler. The Cosmic Ray Antiproton Background for AMS-02. *JCAP*, 1409:051, 2014.

[23] C. Consolandi. Primary Cosmic Ray Proton Flux Measured by AMS-02. In *Proceedings, 33rd International Cosmic Ray Conference (ICRC2013): Rio de Janeiro, Brazil, July 2-9, 2013*, page 1265, 2014. [,1265(2014)].

[24] G. Steigman. Observational tests of antimatter cosmologies. *Ann. Rev. Astron. Astrophys.*, 14:339–372, 1976.

[25] Andrew G. Cohen, A. De Rujula, and S. L. Glashow. A Matter - antimatter universe? *Astrophys. J.*, 495:539–549, 1998.

[26] G. et al. Hinshaw. Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Parameter Results. *APJS*, 208:19, October 2013.

[27] C. L. et al. Bennett. Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Final Maps and Results. *APJS*, 208:20, October 2013.

[28] P. A. R. Ade et al. Planck 2015 results. XIII. Cosmological parameters. *Astron. Astrophys.*, 594:A13, 2016.

[29] J.L. Raaf. Recent proton decay results from super-kamiokande. *Nuclear Physics B - Proceedings Supplements*, 221:387 – 388, 2011. The Proceedings of the 22nd International Conference on Neutrino Physics and Astrophysics.

[30] Gerard 't Hooft. Symmetry Breaking Through Bell-Jackiw Anomalies. *Phys. Rev. Lett.*, 37:8–11, 1976.

[31] Lewis H. Ryder. *Quantum Field Theory*. Cambridge University Press, 1996.

[32] G. Gamow. Expanding universe and the origin of elements. *Phys. Rev.*, 70:572–573, Oct 1946.

[33] Hannu Kurki-Suonio. *Cosmology Lecture Notes*. University of Helsinki.

[34] Andrew G. Cohen and David B. Kaplan. Thermodynamic Generation of the Baryon Asymmetry. *Phys. Lett.*, B199:251–258, 1987.

[35] A. D. Dolgov. Baryogenesis, 30 years after. In *Surveys in high-energy physics. Proceedings, 25th ITEP Winter School, Snegiri, Russia, February, 1997*, 1997.

[36] Antonio De Felice and Shinji Tsujikawa. f(R) theories. *Living Rev. Rel.*, 13:3, 2010.

[37] O. Bertolami and J. Páramos. Mimicking dark matter through a non-minimal gravitational coupling with matter. *JCAP*, 3:009, March 2010.

[38] O. Bertolami, P. Frazao, and J. Paramos. Accelerated expansion from a non-minimal gravitational coupling to matter. *Phys. Rev.*, D81:104046, 2010.

[39] Valerio Faraoni, Edgard Gunzig, and Pasquale Nardone. Conformal transformations in classical gravitational theories and in cosmology. *Fund. Cosmic Phys.*, 20:121, 1999.

[40] Orfeu Bertolami, Christian G. Boehmer, Tiberiu Harko, and Francisco S. N. Lobo. Extra force in f(R) modified theories of gravity. *Phys. Rev.*, D75:104016, 2007.

[41] Luca Amendola and Domenico Tocchini-Valentini. Stationary dark energy: the present universe as a global attractor. *Physical Review D*, 64(4):043509, 2001.

[42] Shin'ichi Nojiri and Sergei D Odintsov. Dark energy and cosmic speed-up from consistent modified gravity. *arXiv preprint hep-th/0412030*, 2004.

[43] Gianluca Allemandi, Andrzej Borowiec, Mauro Francaviglia, and Sergei D Odintsov. Dark energy dominance and cosmic acceleration in first-order formalism. *Physical Review D*, 72(6):063505, 2005.

[44] Tomi Koivisto. A note on covariant conservation of energy–momentum in modified gravities. *Classical and Quantum Gravity*, 23(12):4289, 2006.

[45] O. Bertolami and J. Páramos. Minimal extension of General Relativity: alternative gravity model with non-minimal coupling between matter and curvature. *Int. J. Geom. Meth. Mod. Phys.*, 11:1460003, 2014.

[46] O. Bertolami and J. Páramos. On the non-trivial gravitational coupling to matter. *Class. Quant. Grav.*, 25:245017, 2008.

[47] Yasunori Fujii. Conformal transformation in the scalar-tensor theory applied to the accelerating universe. *Prog. Theor. Phys.*, 118:983–1018, 2007.

[48] Valerio Faraoni and Shahn Nadeau. The (pseudo)issue of the conformal frame revisited. *Phys. Rev.*, D75:023501, 2007.

[49] J. David Brown. Action functionals for relativistic perfect fluids. *Class. Quant. Grav.*, 10:1579–1606, 1993.

[50] Orfeu Bertolami, Francisco S. N. Lobo, and Jorge Paramos. Non-minimum coupling of perfect fluids to curvature. *Phys. Rev.*, D78:064036, 2008.

[51] I. et al. Prigogine. Thermodynamics of cosmological matter creation. *Proceedings of the National Academy of Sciences of the United States of America*, 85(20):74287432, 1988.

[52] Tiberiu Harko. Thermodynamic interpretation of the generalized gravity models with geometry - matter coupling. *Phys. Rev.*, D90(4):044067, 2014.

[53] L. E. Reichl. *A Modern Course in Statistical Physics.* John Wiley and Sons, Inc., 1998.

[54] Hooman Davoudiasl, Ryuichiro Kitano, Graham D. Kribs, Hitoshi Murayama, and Paul J. Steinhardt. Gravitational baryogenesis. *Phys. Rev. Lett.*, 93:201301, 2004.

[55] Gaetano Lambiase and G. Scarpetta. Baryogenesis in f(R): Theories of Gravity. *Phys. Rev.*, D74:087504, 2006.

[56] William H. Kinney, Edward W. Kolb, Alessandro Melchiorri, and Antonio Riotto. Inflation model constraints from the Wilkinson Microwave Anisotropy Probe three-year data. *Phys. Rev.*, D74:023502, 2006.

[57] H. V. Peiris et al. First year Wilkinson Microwave Anisotropy Probe (WMAP) observations: Implications for inflation. *Astrophys. J. Suppl.*, 148:213–231, 2003.

[58] Michael Dine and Alexander Kusenko. The Origin of the matter - antimatter asymmetry. *Rev. Mod. Phys.*, 76:1, 2003.

[59] Mikael Chala, Germano Nardini, and Ivan Sobolev. Unified explanation for dark matter and electroweak baryogenesis with direct detection and gravitational wave signatures. *Phys. Rev.*, D94(5):055006, 2016.

[60] Mingzhe Li, Jun-Qing Xia, Hong Li, and Xinmin Zhang. Cosmological CPT violation, baryo/leptogenesis and CMB polarization. *Phys. Lett.*, B651:357–362, 2007.

[61] Francisco S. N. Lobo, Tiberiu Harko, Jos P. Mimoso, and Diego Pavn. Irreversible matter creation processes through a nonminimal curvature-matter coupling. 2015.

[62] M. O. Calvão, J. A. S. Lima, and I. Waga. On the thermodynamics of matter creation in cosmology. *Physics Letters A*, 162:223–226, February 1992.

[63] Jos Pedro Mimoso and Diego Pavn. Entropy evolution of universes with initial and final de Sitter eras. *Phys. Rev.*, D87(4):047302, 2013.

[64] Mark Thomson. *Modern Particle Physics*. Cambridge University Press, 2013.

[65] David Griffiths. *Introduction to Elementary Particles*. Wiley-VCH, 2008.

Appendix A

Neutron Beta-Decay

The vector-axial nature of the weak charged-current interaction gives a vertex factor [64]

$$\mathcal{K}^\mu \to \frac{ig_\omega}{2\sqrt{2}}\gamma^\mu(1-\gamma^5) \ , \tag{A.0.1}$$

at both vertices in the Feynman diagrams involving processes such as (3.5.83). Due to the large mass of the W boson, the weak force propagator can be approximated (providing the transferred momentum satisfies $q \ll M_W^2$) by

$$\mathcal{D}_{\mu\nu} \to -i\frac{g_{\mu\nu}}{M_W^2} \ . \tag{A.0.2}$$

In this type of interaction, we are dealing with the interchange of protons and neutrons, which are composite particles; so, to be more rigorous, we should include in the theory the presence of spectator quarks (one u and one d), otherwise the nucleon's internal structure is not resolved in the theory. This is accomplished by requiring both the vector and axial portions of the quark-quark vertex factor to assume the more general form

$$\mathcal{K}'^\mu \to \frac{ig_\omega}{2\sqrt{2}}\gamma^\mu(g_V - g_A\gamma^5) \ , \tag{A.0.3}$$

where g_V and g_A are constants (they only change the magnitudes of the vector and axial-vector couplings). Experimental data give [65] $g_A \approx 1.27$ and $g_V = 1.00$; the vector weak charge is not modified by the strong interactions within the nucleon.

Given these structural elements, one can compute the mean neutron lifetime. Inserting the functional form of the propagator and the vertex factors, the matrix element for the composite neutron follows as

$$i\mathcal{M} = \frac{g_\omega^2}{8M_W^2}[\overline{u}(p)\gamma^\mu(1-g_A\gamma^5)u(n)]\cdot[\overline{u}(e)\gamma_\mu(1-\gamma^5)v(\nu_e)] \ . \tag{A.0.4}$$

Recalling Feynman rules, we know that to each particle entering the diagram one associates a spinor $u_s(\boldsymbol{p})$ to the respective external line; if, on the contrary, the particle is exiting the diagram, one associates $\overline{u}_s(\boldsymbol{p})$. In the case of an antiparticle entering the diagram, it is represented by $\overline{v}_s(\boldsymbol{p})$; or, if it's exiting, $v_s(\boldsymbol{p})$. Here $\boldsymbol{p}$ is the momentum and $s = 1, 2$ refers to the two spin states. The Dirac spinors form a complete base, in the sense that

$$\sum_s u_s \overline{u}_s = (\not{p} + m) \ , \quad \sum_s v_s \overline{v}_s = (\not{p} - m) \ . \tag{A.0.5}$$

Now, to evaluate the square of (A.0.4), averaged over all spin states, we must deal with a general quantity of the form

$$G \equiv [\overline{u}(a)\Gamma_1 u(b)][\overline{u}(a)\Gamma_2 u(b)]^* \ , \tag{A.0.6}$$

where a and b denote the spins and momenta and Γ_1 and Γ_2 are combinations of 4×4 matrices. First, note that

$$[\overline{u}(a)\Gamma_2 u(b)]^* = [\overline{u}(a)^\dagger \gamma^0 \Gamma_2 u(b)]^\dagger = \overline{u}(b)^\dagger \Gamma_2^\dagger \gamma^{0\dagger} u(a) = \overline{u}(b)^\dagger \gamma^0 \gamma^0 \Gamma_2^\dagger \gamma^{0\dagger} u(a) = \overline{u}(b) \overline{\Gamma}_2 u(a) \ , \tag{A.0.7}$$

where

$$\overline{\Gamma}_2 \equiv \gamma^0 \Gamma_2^\dagger \gamma^0 \ , \tag{A.0.8}$$

and we used the facts that $\gamma^{0\dagger} = \gamma^0$ and $\left(\gamma^0\right)^2 = 1$. Now we start summing the spin orientations of particle b. We use the completeness relation (A.0.5) to get

$$\begin{aligned} \sum_{s_b} G &= \overline{u}(a)\Gamma_1 \left\{ \sum_{s_b} u_{s_b}(p_b) \overline{u}_{s_b}(p_b) \right\} \overline{\Gamma}_2 u(a) \\ &= \overline{u}(a)\Gamma_1 (\not{p}_b + m_b) \overline{\Gamma}_2 u(a) \ . \end{aligned} \tag{A.0.9}$$

Calling $Q \equiv \Gamma_1(\not{p}_b + m_b)\overline{\Gamma}_2$, we do the same summation for particle a:

$$\begin{aligned} \sum_{s_a} \overline{u}_{s_a}(p_a)_i Q_{ij} u_{s_a}(p_a)_j &= Q_{ij} \left\{ \sum_{s_a} u_{s_a}(p_a) \overline{u}_{s_a}(p_a) \right\}_{ji} \\ &= Q_{ij}(\not{p}_a + m_a)_{ji} = [Q(\not{p}_a + m_a)]_{ii} = \mathrm{Tr}[Q(\not{p}_a + m_a)] \ . \end{aligned} \tag{A.0.10}$$

So, in conclusion, we get:

$$\sum_{\text{all spins}} [\overline{u}(a)\Gamma_1 u(b)][\overline{u}(a)\Gamma_2 u(b)]^* = \mathrm{Tr}[\Gamma_1(\not{p}_b + m_b)\overline{\Gamma}_2(\not{p}_a + m_a)] \ . \tag{A.0.11}$$

If either u is replaced by a v-spinor, the corresponding mass on the r.h.s. switches sign.

Next, we apply this *trick* to calculate explicitly the square of the $\mathcal{M}$-matrix in (A.0.4):

$$\begin{aligned} \langle |\mathcal{M}|^2 \rangle = \frac{1}{2} \cdot \frac{g_\omega^2}{64 M_W^4} &\mathrm{Tr}[\gamma^\mu(1 - g_A\gamma^5)(\not{p}_p + m_p)\gamma^\alpha(1 - g_A\gamma^5)(\not{p}_n + m_n)] \\ &\times \mathrm{Tr}[\gamma_\mu(1 - \gamma^5)(\not{p}_e + m_e)\gamma_\alpha(1 - \gamma^5)\not{p}_{\nu_e}] \end{aligned} \tag{A.0.12}$$

Note that $\Gamma_2 = \gamma^\mu\gamma^5$; hence $\overline{\Gamma}_2 = \gamma^0(\gamma^\mu\gamma^5)^\dagger\gamma^0 = \gamma^0\gamma^5\gamma^0\gamma^0\gamma^{\mu\dagger}\gamma^0 = -\gamma^5\gamma^\mu = \gamma^\mu\gamma^5$. The factor 1/2 is included when averaging over the initial spins, in this case, two spin states of the neutron.

The second trace **Tr[2]** in this equation is calculated considering the following elements:

$$\begin{aligned} \mathrm{Tr}[\not{p}_\nu \gamma_\mu \not{p}_e \gamma_\alpha] &= p_\nu^\rho p_e^\sigma \mathrm{Tr}[\gamma_\rho, \gamma_\mu, \gamma_\sigma, \gamma_\alpha] \\ &= 4 p_\nu^\rho p_e^\sigma (g_{\rho\mu} g_{\sigma\alpha} - g_{\rho\sigma} g_{\mu\alpha} + g_{\rho\alpha} g_{\mu\sigma}) \\ &= 4(p_{\nu,\mu} p_{e,\alpha} + p_{\nu,\alpha} p_{e,\mu} - g_{\mu\alpha} p_\nu \cdot p_e) \ ; \end{aligned} \tag{A.0.13}$$

$$\begin{aligned}\mathrm{Tr}[-\not{p}_\nu\gamma_\mu\gamma_5\not{p}_e\gamma_\alpha] &= -p_\nu^\rho p_e^\sigma \mathrm{Tr}[\gamma_\rho\gamma_\mu\gamma_5\gamma_\sigma\gamma_\alpha] \\ &= -p_\nu^\rho p_e^\sigma \mathrm{Tr}[\gamma_5\gamma_\sigma\gamma_\alpha\gamma_\rho\gamma_\mu] \\ &= -4i\epsilon_{\sigma\alpha\rho\mu} p_\nu^\rho p_e^\sigma \ .\end{aligned} \tag{A.0.14}$$

Noting that $\mathrm{Tr}[\not{p}_\nu\gamma_\mu\gamma_5\not{p}_e\gamma_\alpha\gamma_5] = p_\nu^\rho p_e^\sigma \mathrm{Tr}[\gamma_\rho\gamma_\mu\gamma_5\gamma_\sigma\gamma_\alpha\gamma_5] = p_\nu^\rho p_e^\sigma \mathrm{Tr}[\gamma_\rho\gamma_\mu\gamma_5\gamma_5\gamma_\sigma\gamma_\alpha]$ since γ_5 anticommutes with every γ_μ ($\mu = 0,1,2,3$). Then, using $(\gamma_5)^2 = 1$, we conclude this element is the same as (A.0.13). Similarly, the term $\mathrm{Tr}[-\not{p}_\nu\gamma_\mu\not{p}_e\gamma_\alpha\gamma_5] = p_\nu^\rho p_e^\sigma \mathrm{Tr}[-\gamma_\rho\gamma_\mu\gamma_\sigma\gamma_\alpha\gamma_5] = p_\nu^\rho p_e^\sigma \mathrm{Tr}[\gamma_\rho\gamma_\mu\gamma_\sigma\gamma_5\gamma_\alpha] = p_\nu^\rho p_e^\sigma \mathrm{Tr}[-\gamma_\rho\gamma_\mu\gamma_5\gamma_\sigma\gamma_\alpha]$ equals the term in (A.0.14). Hence

$$\mathbf{Tr[2]} = 8[p_{\nu,\mu}p_{e,\alpha} + p_{\nu,\alpha}p_{e,\mu} - g_{\mu\alpha}(p_\nu \cdot p_e) - i\epsilon_{\sigma\alpha\rho\mu} p_\nu^\rho p_e^\sigma] \ , \tag{A.0.15}$$

where we have neglected the electron mass. The same elements appear in the first trace **Tr[1]** of (A.0.12). Now we are in a position to calculate **Tr[1]**×**Tr[2]**.

Let's first work out the elements independent of the masses. The zero-order term in g_A reads

$$(\mathbf{Tr[1]} \times \mathbf{Tr[2]})_0 = 64[(p_n \cdot p_\nu)(p_p \cdot p_e) + (p_n \cdot p_e)(p_p \cdot p_\nu)] \ , \tag{A.0.16}$$

contracting all the terms and noting that the sum $(p_n^\mu p_p^\alpha + p_n^\alpha p_p^\mu) p_\nu^\rho p_e^\sigma \epsilon_{\sigma\alpha\rho\mu}$ vanishes, because the term in parentheses is symmetric under the interchange of $\mu \leftrightarrow \alpha$ while the Levi-Civita is anti-symmetric ($\epsilon_{\sigma\mu\rho\alpha} = \epsilon_{\sigma\rho\alpha\mu} = -\epsilon_{\sigma\alpha\rho\mu}$).

Considering the same argument, the first-order term in g_A reduces to

$$\begin{aligned}(\mathbf{Tr[1]} \times \mathbf{Tr[2]})_1 &= 64 g_A[-i\epsilon^{\sigma'\alpha\rho'\mu} p_{n,\rho'} p_{p,\sigma'}][-i\epsilon_{\sigma\alpha\rho\mu} p_\nu^\rho p_e^\sigma] \\ &= -64 g_A[p_{n,\rho'} p_{p,\sigma'} p_\nu^\rho p_e^\sigma \epsilon_{\alpha\mu\sigma\rho} \epsilon^{\alpha\mu\sigma' rho'}] \\ &= 128 g_A[p_{n,\rho'} p_{p,\sigma'} p_\nu^\rho p_e^\sigma (\delta_\sigma^{\sigma'}\delta_\rho^{\rho'} - \delta_\rho^{\sigma'}\delta_\sigma^{\rho'})] \\ &= 128 g_A[(p_p \cdot p_e)(p_n \cdot p_\nu) - (p_p \cdot p_\nu)(p_n \cdot p_e)] \ ,\end{aligned} \tag{A.0.17}$$

where we have use the identity $\epsilon^{\mu\nu\lambda\sigma}\epsilon_{\mu\nu\kappa\tau} = -2(\delta_\kappa^\lambda\delta_\tau^\sigma - \delta_\tau^\lambda\delta_\kappa^\sigma)$.

The second-order term in g_A appearing in **Tr[1]** is exactly the same as the zero-order term, since $\mathrm{Tr}[\gamma^\rho\gamma^\mu\gamma^5\gamma^\sigma\gamma^\alpha\gamma^5] = \mathrm{Tr}[\gamma^\rho\gamma^\mu\gamma^\sigma\gamma^\alpha]$, hence

$$(\mathbf{Tr[1]} \times \mathbf{Tr[2]})_2 = (\mathbf{Tr[1]} \times \mathbf{Tr[2]})_0 \ . \tag{A.0.18}$$

Now considering only the terms proportional to $m_n m_p$, the first trace reads

$$m_n m_p(\mathrm{Tr}[\gamma^\mu\gamma^\alpha] - 2g_A \mathrm{Tr}[\gamma^\mu\gamma^5\gamma^\alpha] + g_A^2 \mathrm{Tr}[\gamma^\mu\gamma^5\gamma^\alpha\gamma^5]) \ . \tag{A.0.19}$$

The trace of two γ-matrices is obtained using the fundamental property $\{\gamma^\mu, \gamma^\nu\} = 2g^{\mu\nu}$ and inserting γ^5-matrices which satisfy $(\gamma^5)^2 = 1$. First, consider the trace of the product of two γ-matrices, $\mathrm{Tr}[\gamma^\mu\gamma^\alpha] = \mathrm{Tr}[2g^{\mu\alpha}\mathbf{1} - \gamma^\alpha\gamma^\mu] = 8g^{\mu\alpha} - \mathrm{Tr}[\gamma^\mu\gamma^\alpha] \Rightarrow \mathrm{Tr}[\gamma^\mu\gamma^\alpha] = 4g^{\mu\alpha}$, where we have used the cyclic property of the trace. The first-order term in g_A is zero, since $\mathrm{Tr}[\gamma^\mu\gamma^5\gamma^\alpha] = 0$; and the second-order term is simplified to $\mathrm{Tr}[-\gamma^\mu\gamma^\alpha] = -4g^{\mu\alpha}$.

The contraction of this term with **Tr[2]** gives

$$\begin{aligned} & m_n m_p[(1-g_A^2)4g^{\mu\alpha}] \times 8[p_{\nu,\mu}p_{e,\alpha} + p_{\nu,\alpha}p_{e,\mu} - g_{\mu\alpha}(p_\nu \cdot p_e) - i\epsilon_{\sigma\alpha\rho\mu}p_\nu^\rho p_e^\sigma] \\ & = 32 m_n m_p (1-g_A^2)[(p_\nu \cdot p_e)(2 - g^{\mu\alpha}g_{\mu\alpha})] \\ & = 64 m_n m_p (g_A^2 - 1)(p_\nu \cdot p_e) \ . \end{aligned} \tag{A.0.20}$$

Hence, the matrix $\langle|\mathcal{M}|^2\rangle$ can be written as the sum of three terms:

$$\langle|\mathcal{M}_1|^2\rangle = \frac{1}{2}(1+g_A)^2\left[\frac{g_\omega}{M_W}\right]^4 (p_n \cdot p_\nu)(p_p \cdot p_e) \ ; \tag{A.0.21}$$

$$\langle|\mathcal{M}_2|^2\rangle = \frac{1}{2}(1-g_A)^2\left[\frac{g_\omega}{M_W}\right]^4 (p_n \cdot p_e)(p_\nu \cdot p_p) \ ; \tag{A.0.22}$$

$$\langle|\mathcal{M}_3|^2\rangle = \frac{1}{2}(g_A^2-1)\left[\frac{g_\omega}{M_W}\right]^4 m_n m_p (p_\nu \cdot p_e) \ . \tag{A.0.23}$$

The next step is to evaluate the dot products in the above expressions. We assume a referential frame where the neutron is initially at rest, so $p_n = (m_n, \vec{0})$; then:

$$(p_n \cdot p_\nu) = E_\nu m_n \tag{A.0.24}$$

and, because of the conservation law $p_p + p_e = p_n + p_\nu$, we can write:

$$(p_p + p_e)^2 = m_p^2 + m_e^2 + 2(p_p \cdot p_e) = (p_n - p_\nu)^2 = m_n^2 - 2(p_n \cdot p_\nu) \ . \tag{A.0.25}$$

Making use of equation (A.0.24), we get:

$$(p_p \cdot p_e) = \frac{1}{2}(m_n^2 - m_p^2 - m_e^2 - 2E_\nu m_n) \ , \tag{A.0.26}$$

so that

$$(p_n \cdot p_\nu)(p_p \cdot p_e) = \frac{1}{2} m_n E_\nu (m_n^2 - m_p^2 - m_e^2 - 2E_\nu m_n) \ . \tag{A.0.27}$$

Similarly,

$$(p_n \cdot p_e)(p_\nu \cdot p_p) = \frac{1}{2} m_n E_e (m_n^2 - m_p^2 + m_e^2 - 2m_n E_e) \ ; \tag{A.0.28}$$

and

$$(p_\nu \cdot p_e) = E_\nu E_e - \boldsymbol{p}_\nu \cdot \boldsymbol{p_e} \ . \tag{A.0.29}$$

The decay rate calculation proceeds as usual:

$$\begin{aligned} d\Gamma = \frac{\langle|\mathcal{M}|^2\rangle}{2m_n}\left(\frac{d^3\boldsymbol{p}_\nu}{(2\pi)^3 2|\boldsymbol{p}_\nu|}\right)\left(\frac{d^3\boldsymbol{p}_p}{(2\pi)^3 2\sqrt{\boldsymbol{p}_p^2 + m_p^2}}\right)\left(\frac{d^3\boldsymbol{p}_e}{(2\pi)^3 2\sqrt{\boldsymbol{p}_e^2 + m_e^2}}\right) \\ \times (2\pi)^4 \delta^4(p_n - p_\nu - p_p - p_e) \ . \end{aligned} \tag{A.0.30}$$

The $\boldsymbol{p}_p$ integral gives

$$d\Gamma = \frac{\langle|\mathcal{M}|^2\rangle}{16(2\pi)^5 m_n}\frac{d^3\boldsymbol{p}_\nu d^3\boldsymbol{p}_e}{|\boldsymbol{p}_\nu| u \sqrt{\boldsymbol{p}_e^2 + m_e^2}}\delta\left(m_n - |\boldsymbol{p}_\nu| - u - \sqrt{\boldsymbol{p}_e^2 + m_e^2}\right) \ , \tag{A.0.31}$$

where $u \equiv \sqrt{(\boldsymbol{p}_\nu + \boldsymbol{p}_e)^2 + m_p^2}$. Note that with this integration, $\boldsymbol{p}_p$ is replaced by $(\boldsymbol{p}_\nu + \boldsymbol{p}_e)$ due to the $\delta^3(\boldsymbol{p}_n - \boldsymbol{p}_\nu - \boldsymbol{p}_p - \boldsymbol{p}_e)$ factor ($\boldsymbol{p}_n = \vec{0}$ in our reference frame) appearing in the initial integrand.

Performing the $\boldsymbol{p}_\nu$ integral, we use

$$d^3\boldsymbol{p}_\nu = |\boldsymbol{p}_\nu|^2 d|\boldsymbol{p}_\nu| \sin\theta d\theta d\phi \ . \tag{A.0.32}$$

Since $\boldsymbol{p}_e$ is fixed as we work out the $\boldsymbol{p}_\nu$ integral, we can choose $\boldsymbol{p}_e$ to align with the z axis, so that the angle between these two vectors is just θ. Then:

$$u^2 = |\boldsymbol{p}_\nu|^2 + |\boldsymbol{p}_e|^2 + 2|\boldsymbol{p}_\nu||\boldsymbol{p}_e|\cos\theta + m_p^2 \ . \tag{A.0.33}$$

Next, we integrate the angular part of $\boldsymbol{p}_\nu$, using that $udu = -|\boldsymbol{p}_\nu||\boldsymbol{p}_e|\sin\theta d\theta$ for a fixed modulus. Noting that

$$\frac{d^3\boldsymbol{p}_\nu}{|\boldsymbol{p}_\nu|u} = \frac{|\boldsymbol{p}_\nu|d|\boldsymbol{p}_\nu|\sin\theta d\theta d\phi}{u} = \frac{-dud\phi}{|\boldsymbol{p}_e|}d|\boldsymbol{p}_\nu| \ , \tag{A.0.34}$$

and integrating ϕ and θ (or rather u) in equation (A.0.31), we obtain:

$$d\Gamma = \frac{\langle|\mathcal{M}|^2\rangle}{(4\pi)^4 m_n}\frac{d^3\boldsymbol{p}_e}{|\boldsymbol{p}_e|\sqrt{\boldsymbol{p}_e^2 + m_e^2}}d|\boldsymbol{p}_\nu|I \ , \tag{A.0.35}$$

where

$$I \equiv \int_{u_-}^{u_+} \delta\left(m_n - |\boldsymbol{p}_\nu| - \sqrt{\boldsymbol{p}_e^2 + m_e^2} - u\right) du \ . \tag{A.0.36}$$

The minus sign in (A.0.34) just changes the limits of integration of θ, so the lower limit becomes $\theta_- = \pi$ and the upper one $\theta_+ = 0$; in terms of the u-variable, this is equivalent to

$$u_\pm = \sqrt{(|\boldsymbol{p}_\nu| \pm |\boldsymbol{p}_e|)^2 + m_p^2} \ . \tag{A.0.37}$$

Equation (A.0.36) defines the range of the $|\boldsymbol{p}_\nu|$ integral: I is non-vanishing, equal to unity, if and only if $u_- < m_n - |\boldsymbol{p}_\nu| - \sqrt{\boldsymbol{p}_e^2 + m_e^2} < u_+$; thus, $|E_\nu|$ takes only values obeying

$$(E_\nu - |\boldsymbol{p}_e|)^2 + m_p^2 < F(E_\nu) < (E_\nu + |\boldsymbol{p}_e|)^2 + m_p^2 \ , \tag{A.0.38}$$

where

$$F(E_\nu) = m_n^2 + E_\nu^2 + (|\boldsymbol{p}_e|^2 + m_e^2) + 2E_\nu\sqrt{|\boldsymbol{p}_e|^2 + m_e^2} - 2m_n(E_\nu + \sqrt{|\boldsymbol{p}_e|^2 + m_e^2}) \ . \tag{A.0.39}$$

It is straightforward then to obtain the maximum and minimum values of E_ν; for example, working out the l.h.s. inequality, we get

$$\begin{aligned} & -2E_\nu|\boldsymbol{p}_e| + m_p^2 < m_n^2 + m_e^2 + 2E_\nu\sqrt{|\boldsymbol{p}_e|^2 + m_e^2} - 2m_nE_\nu - 2m_n\sqrt{|\boldsymbol{p}_e|^2 + m_e^2} \\ & \Leftrightarrow -2E_\nu(|\boldsymbol{p}_e| - \sqrt{|\boldsymbol{p}_e|^2 + m_e^2} + m_n) < m_n^2 - m_p^2 + m_e^2 - 2m_n\sqrt{|\boldsymbol{p}_e|^2 + m_e^2} \\ & \Leftrightarrow E_\nu > \frac{\frac{1}{2}(m_n^2 - m_p^2 + m_e^2) - m_n\sqrt{|\boldsymbol{p}_e|^2 + m_e^2}}{m_n - \sqrt{|\boldsymbol{p}_e|^2 + m_e^2} + |\boldsymbol{p}_e|} \ , \end{aligned} \tag{A.0.40}$$

which defines the minimum value for the energy of the neutrino in the present decay. The maximum is obtained working out the r.h.s. of (A.0.38). Gathering the results for the maximum and the minimum, we obtain:

$$p_{\pm} = \frac{\frac{1}{2}(m_n^2 - m_p^2 + m_e^2) - m_n\sqrt{|\boldsymbol{p}_e|^2 + m_e^2}}{m_n - \sqrt{|\boldsymbol{p}_e|^2 + m_e^2} \mp |\boldsymbol{p}_e|} \ . \tag{A.0.41}$$

We are now able to carry out the E_ν (or $|\boldsymbol{p}_\nu|$) integral in each $\langle|\mathcal{M}_i|^2\rangle$ (equations A.0.21, A.0.22, A.0.23). Starting with $\langle|\mathcal{M}_1|^2\rangle$, the $|\boldsymbol{p}_\nu|$ integral reads

$$\int_{p_-}^{p^+} |\boldsymbol{p}_\nu|(m_n^2 - m_p^2 - m_e^2 - 2m_n|\boldsymbol{p}_\nu|)d|\boldsymbol{p}_\nu| \equiv J_1 \tag{A.0.42}$$

Remember that in the case of the neutrino, writing $|\boldsymbol{p}_\nu|$ or E_ν is the same. Additionally, since

$$d^3\boldsymbol{p}_e = 4\pi|\boldsymbol{p}_e|^2 d|\boldsymbol{p}_e| \ , \tag{A.0.43}$$

and recalling the relation $m_e^2 = p_e^2 = E_e^2 - |\boldsymbol{p}_e|^2 \Rightarrow |\boldsymbol{p}_e| = \sqrt{E_e^2 - m_e^2}$, which also implies $d|\boldsymbol{p}_e| = \frac{E_e}{|\boldsymbol{p}_e|}dE_e$, we can write the volume element that appears in (A.0.35) as

$$\frac{d^3\boldsymbol{p}_e}{|\boldsymbol{p}_e|\sqrt{\boldsymbol{p}_e^2 + m_e^2}} = 4\pi dE_e \ . \tag{A.0.44}$$

Hence, we conclude that

$$\frac{d\Gamma_1}{dE_e} = \frac{1}{4}\frac{(1+g_A)^2}{(4\pi)^3}\left[\frac{g_\omega}{M_W}\right]^4 J_1(E_e) \ . \tag{A.0.45}$$

Although the integral is trivial to calculate, $J(E_e)$ is not a simple function to deal with:

$$J_1(E_e) = \frac{1}{2}(m_n^2 - m_p^2 - m_e^2)(p_+^2 - p_-^2) - \frac{2m_n}{3}(p_+^3 - p_-^3) \ . \tag{A.0.46}$$

This can be, however, approximated noting we are dealing with four small numbers:

$$\epsilon \equiv \frac{m_n - m_p}{m_n} \ , \quad \delta \equiv \frac{m_e}{m_n} \ , \quad \eta \equiv \frac{E_e}{m_n} \ , \quad \phi \equiv \frac{|\boldsymbol{p}_e|}{m_n} \ . \tag{A.0.47}$$

In the final result, we can neglect the electron mass in equation (A.0.41). Also, writing $p_\pm$ in terms of these parameters, we get

$$p_+ = \frac{(m_n^2\epsilon - m_n^2\eta)}{m_n(1-\eta-\phi)} \ ; \tag{A.0.48}$$

whereas

$$p_- = \frac{(m_n^2\epsilon - m_n^2\eta)}{m_n(1-\eta+\phi)} \ , \tag{A.0.49}$$

where we have used $(m_n + m_p) \approx 2m_n$. Expanding to the lowest order the following terms:

$$\left[\frac{1}{(1-\eta-\phi)^2} - \frac{1}{(1-\eta+\phi)^2}\right] \approx [1 + 2(\eta+\phi) - 1 - 2(\eta-\phi)] = 4\phi \ ; \tag{A.0.50}$$

$$\left[\frac{1}{(1-\eta-\phi)^3} - \frac{1}{(1-\eta+\phi)^3}\right] \approx 1 + 3(\eta+\phi) - [1 + 3(\eta-\phi)] = 6\phi \ ; \tag{A.0.51}$$

we obtain

$$J_1 \approx 4m_n^4\phi\epsilon(\epsilon-\eta)^2\phi - 4m_n^4\phi(\epsilon-\eta)^3 = 4m_n^4\eta\phi(\epsilon-\eta)^2 = 4E_e|\boldsymbol{p}_e|(m_n - m_p - E_e)^2 \ . \tag{A.0.52}$$

Next, let us work out the $|\boldsymbol{p}_\nu|$ integral in $\langle|\mathcal{M}_2|^2\rangle$. In this case, the dot product does not depend on E_ν. We need to calculate the element

$$E_e(m_n^2 - m_p^2 + m_e^2 - 2m_n E_e)\int_{p_-}^{p^+} d|\boldsymbol{p}_\nu| \equiv J_2 \ , \tag{A.0.53}$$

or, in terms of the small parameters we defined,

$$J_2(E_e) = m_n\eta(2m_n^2\epsilon - 2m_n^2\eta)(p_+ - p_-) \ . \tag{A.0.54}$$

Now, expanding the denominator of the sum $(p_+ - p_-)$:

$$\left[\frac{1}{1-\eta-\phi} - \frac{1}{1-\eta+\phi}\right] \approx 1 + (\eta+\phi) - [1 + (\eta-\phi)] = 2\phi \ , \tag{A.0.55}$$

we conclude that

$$J_2 \approx m_n^4[2\eta(\epsilon-\eta)][2\phi(\epsilon-\eta)] = 4m_n^4\eta\phi(\epsilon-\eta)^2 = 4E_e|\boldsymbol{p}_e|(m_n - m_p - E_e)^2 \ . \tag{A.0.56}$$

The dot product in $\langle|\mathcal{M}_2|\rangle$ shows no dependence on the angular part, so again the $\boldsymbol{p}_e$ integral makes a (4π) factor come out, leading to

$$\frac{d\Gamma_2}{dE_e} = \frac{1}{4}\frac{(1-g_A)^2}{(4\pi)^3}\left[\frac{g_\omega}{M_W}\right]^4 J_2(E_e) \ . \tag{A.0.57}$$

The last term $\langle|\mathcal{M}_3|\rangle$ depends not only on E_ν, but also on the angle $\theta_{e\nu}$. The dot product that appears in this case can be recast in the form

$$(p_\nu \cdot p_e) = E_\nu E_e(1 - \cos\theta_{e\nu}) \ , \tag{A.0.58}$$

if we again ignore the mass of the electron. This element has now a non-trivial angular part, which one can argue integrates to zero: looking back into equation (A.0.30), we could have started by solving the angular part of the integration in $\boldsymbol{p}_\nu$, that is $\int\int \cos\theta_{e\nu}\sin\theta d\theta d\phi$. Again, when integrating over one variable, the others act as constants; in particular, we can choose the direction of $\boldsymbol{p}_e$ to be the same as the z-axis, so that $\theta_{e\nu} = \theta$. The delta function fixes the direction of $\boldsymbol{p}_p$ to be $(-\boldsymbol{p}_\nu/|\boldsymbol{p}_\nu| - \hat{z})$, so there is no problem with the conservation of momentum. Then, with these considerations in mind, we simply need to calculate the integral over θ, which is equal to zero. The first term in (A.0.58) introduces a factor of 4π since there is no angular dependence. Then, performing the integral on $|\boldsymbol{p}_\nu|$ as before, we get

$$E_e\int_{p_-}^{p^+} |\boldsymbol{p}_\nu| d\boldsymbol{p}_\nu \equiv J_3 \ , \tag{A.0.59}$$

so that

$$J_3(E_e) = \frac{m_n\eta}{2}(p_+^2 - p_-^2) \ . \tag{A.0.60}$$

Expanding to the lowest order, making use of equation (A.0.50), we obtain:

$$J_3 \approx 2m_n^3\eta\phi(\epsilon-\eta)^2 = \frac{2}{m_n}E_e|\boldsymbol{p}_e|(m_n - m_p - E_e)^2 \ . \tag{A.0.61}$$

Hence,

$$\frac{d\Gamma_3}{dE_e} = \frac{1}{2}\frac{(g_A^2 - 1)}{(4\pi)^3}\left[\frac{g_\omega}{M_W}\right]^4 m_n J_3(E_e) \ , \tag{A.0.62}$$

where again we use the approximation $m_p \approx m_n$.

Finally, summing all the contributions, we obtain the total (differential) decay rate, $\Gamma = \Gamma_1 + \Gamma_2 + \Gamma_3$:

$$\frac{d\Gamma}{dE_e} = \frac{1 + 3g_A^2}{64\pi^3}\left[\frac{g_\omega}{M_W}\right]^4 E_e\sqrt{E_e^2 - m_e^2}(m_n - m_p - E_e)^2 \ . \tag{A.0.63}$$

We can express the coupling factor in terms of the Fermi constant,

$$G_F \equiv \frac{\sqrt{2}}{8}\left[\frac{g_\omega}{M_W}\right]^2 \ , \tag{A.0.64}$$

whose numerical value follows from experimental measurements (mostly of the muon decay rate) [65]:

$$G_F^{\exp} = 1.16637(1) \times 10^{-5}\ \mathrm{GeV}^{-2} \ . \tag{A.0.65}$$

Integrating over electron energies and defining the integral in terms of the dimensionless quantities $q = (m_n - m_p)/m_e = Q/m_e$ and $\xi = E_e/m_e$, we end up with the numerical factor:

$$\lambda_0 \equiv \int_1^q d\xi\xi\sqrt{\xi^2 - 1}(\xi - q)^2 \approx 1.636 \ , \tag{A.0.66}$$

where the electron energy range up to about $E_{max} \approx m_n - m_p$ (as a plot of the decay rate A.0.63 vs electron energy would easily evidentiate). Our final result for the neutron decay rate (or inverse of the lifetime) is thus

$$\tau_n^{-1} \equiv \Gamma_{n\to pe\nu} = \frac{G_F^2}{2\pi^3}(1 + 3g_A^2)m_e^5\lambda_0 \ . \tag{A.0.67}$$

Appendix B

Commutator Function

We now give a definition of the Delta function, obtained through the commutation relations of the various fields. We consider the case of a scalar field.

As usual, an arbitrary solution of the Klein-Gordon equation can be expanded as a Fourier integral over plane-wave solutions:

$$\phi(x) = \int \frac{d^3k}{\sqrt{(2\pi)^3 2\omega_k}} [a_+(k)e^{-ik\cdot x} + a_-^\dagger(k)e^{ik\cdot x}] \ , \tag{B.0.1}$$

and

$$\phi^*(x) = \int \frac{d^3k}{\sqrt{(2\pi)^3 2\omega_k}} [a_+^\dagger(k)e^{ik\cdot x} + a_-(k)e^{-ik\cdot x}] \ . \tag{B.0.2}$$

where $\omega_k = +\sqrt{k^2 + m^2}$.

Hence, the commutation function reads:

$$\begin{aligned}
[\phi(x), \phi^*(y)] &= \int \frac{d^3k d^3k'}{(2\pi)^3 \sqrt{2\omega_k 2\omega_{k'}}} ([a_+(k), a_+^\dagger(k')]e^{-ik\cdot x + ik'\cdot y} + [a_-^\dagger(k), a_-(k')]e^{ik\cdot x - ik'\cdot y}) \\
&= \int \frac{d^3k}{(2\pi)^3 2\omega_k} (e^{-ik\cdot(x-y)} - e^{ik\cdot(x-y)}) \\
&\equiv i\Delta(x-y) \ ,
\end{aligned} \tag{B.0.3}$$

so that $\Delta(x)$ is a Lorentz invariant, real function. In its derivation, we used the facts that $[a(k), a^\dagger(k')] = \delta^3(\boldsymbol{k} - \boldsymbol{k'})$ and $[a(k), a(k')] = [a^\dagger(k), a^\dagger(k')] = 0$. As required by the definition in terms of the commutator on the left-hand side of this equation, Δ is a solution of the free Klein-Gordon equation and is an odd function of its argument.

The anticommutation or commutation relations for the other fields are easily obtained, taking into account the general relations for the creation and annihilation operators (these can be found, for example, in Drell's *Relativistic Quantum Fields*). In what concerns the topic of discrete operations, it is much more interesting to infer the symmetry properties of the Δ function. For reasons of transparency, it is better to realize that (B.0.3) is equivalent to the form

$$i\Delta(x-y) = \int \frac{d^3k}{(2\pi)^3 2\omega_k} (e^{-i\omega_k(x_0-y_0) + i\boldsymbol{k}\cdot(\boldsymbol{x}-\boldsymbol{y})} - e^{i\omega_k(x_0-y_0) - i\boldsymbol{k}\cdot(\boldsymbol{x}-\boldsymbol{y})}) \ , \tag{B.0.4}$$

We now change the variable $\overrightarrow{k}$ to $-\overrightarrow{k}$ in the second integral. Noting that $\omega_k = \sqrt{k^2 + m^2}$ is unchanged, we can put the spatial part of both the exponential functions (that are now equal) in evidence, to get:

$$i\Delta(x-y) = -\frac{i}{(2\pi)^3} \int \frac{d^3k}{\omega_k} e^{i\boldsymbol{k}\cdot(\boldsymbol{x}-\boldsymbol{y})} \sin\omega_k(x_0 - y_0) \tag{B.0.5}$$

From these considerations, it follows that $\Delta(\boldsymbol{r}, t) = \Delta(-\boldsymbol{r}, t) = -\Delta(\boldsymbol{r}, -t)$, with $\boldsymbol{r} = \boldsymbol{x} - \boldsymbol{y}$ and $t = x_0 - y_0$. When we change $\boldsymbol{r} \to -\boldsymbol{r}$, the transformed function looks exactly the same as when we change $\boldsymbol{k} \to -\boldsymbol{k}$ and, like before, this does not change the integral. On the other hand, under the transformation $t \to -t$, we get a minus sign, because the sine function is odd. Then, $\Delta(\boldsymbol{r}, t) = -\Delta(-\boldsymbol{r}, -t)$. This result plays a major role in Lüders derivation of the CPT Theorem.

Appendix C

Dirac Matrices and Spinors

The γ matrices in Dirac equation (2.1.3) satisfy the anticommutation relations (2.1.6). A familiar representation of these matrices is:

$$\gamma^0 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} , \tag{C.0.1}$$

$$\gamma^i = \begin{pmatrix} 0 & \sigma^i \\ -\sigma^i & 0 \end{pmatrix} , \tag{C.0.2}$$

where

$$\sigma^1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} , \quad \sigma^2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} , \quad \sigma^3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{C.0.3}$$

are the usual 2×2 Pauli matrices.

Frequently appearing combinations are

$$\sigma^{\mu\nu} = \frac{i}{2}[\gamma^\mu, \gamma^\nu] \quad \text{and} \quad \gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3 . \tag{C.0.4}$$

In this representation, the components of $\sigma^{\mu\nu}$ are:

$$\sigma^{ij} = \begin{pmatrix} \sigma^k & 0 \\ 0 & \sigma^k \end{pmatrix} , \tag{C.0.5}$$

with $i, j, k = 1, 2, 3$ in cyclic order, and

$$\sigma^{0i} = -\sigma^{i0} = i \begin{pmatrix} 0 & \sigma^i \\ \sigma^i & 0 \end{pmatrix} . \tag{C.0.6}$$

The chirality matrix reads:

$$\gamma^5 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} , \tag{C.0.7}$$

with the property

$$\{\gamma^\mu, \gamma^5\} = 0 , \tag{C.0.8}$$

which — in turn — implies that

$$[\gamma^5, \sigma^{\mu\nu}] = 0 \ . \tag{C.0.9}$$

For the inner product of a γ matrix with an ordinary four-vector, we use the following notation:

$$\gamma_\mu A^\mu \equiv \not{A} = \gamma^0 A^0 - \boldsymbol{\gamma} \cdot \boldsymbol{A}, \tag{C.0.10}$$

consistent with the metric signature $(+, -, -, -)$.